Introduction to Longitudinal Research

The recent introduction of computerised dynamic surveys in Europe and North America has made longitudinal data widely available to students and social researchers. In her book, Elisabetta Ruspini provides a concise and comprehensive introduction to the kinds of issues involved in using longitudinal data for the first time. In particular, she covers:

- the advantages of using longitudinal data
- guidance on the availability of longitudinal datasets in Europe, the US and Canada
- the implications of integrating micro level empirical research with macro level theories of social change
- the choices that need to be made – for example, between using trend, panel and duration data.

Introduction to Longitudinal Data will be essential reading for students and social researchers thinking of using longitudinal datasets at any level of complexity.

Elisabetta Ruspini is a lecturer and researcher based in the Department of Sociology and Social Research at the University of Milan-Bicocca.

Social Research Today
edited by Martin Bulmer

The *Social Research Today* series provides concise and contemporary introductions to significant methodological topics in the social sciences. Covering both quantitative and qualitative methods, this new series features readable and accessible books from some of the leading names in the field and is aimed at students and professional researchers alike. This series also brings together for the first time the best titles from the old *Social Research Today* and *Contemporary Social Research* series edited by Martin Bulmer for UCL Press and Routledge. Other series titles include:

Principles of Research Design in the Social Sciences *by Frank Bechhofer and Lindsay Paterson*

Social Impact Assessment *by Henk Becker*

The Turn to Biographical Methods in Social Science *edited by Prue Chamberlayne, Joanna Bornat and Tom Wengraf*

Quantity and Quality in Social Research *by Alan Bryman*

Research Methods and Organisational Studies *by Alan Bryman*

Field Research: A Sourcebook and Field Manual *by Robert G. Burgess*

In the Field: An Introduction to Field Research *by Robert G. Burgess*

Qualitative Analysis: Thinking, Doing, Writing *by Douglas Ezzy*

Research Design second edition *by Catherine Hakim*

Measuring Health and Medical Outcomes *edited by Crispin Jenkinson*

Methods of Criminological Research *by Victor Jupp*

Information Technology for the Social Scientist *edited by Raymond M. Lee*

An Introduction to the Philosophy of Social Research *by Tim May and Malcolm Williams*

Researching Social and Economic Changes: The Uses of Household Panel Studies *edited by David Rose*

Introduction to Longitudinal Research *by Elisabetta Ruspini*

Surveys in Social Research fifth edition *by David de Vaus*

Researching the Powerful in Education *edited by Geoffrey Walford*

Martin Bulmer *is Professor of Sociology and co-director of the Institute of Social Research at the University of Surrey. He is also Academic Director of the Question Bank in the ESRC Centre for Applied Social Surveys, London.*

Introduction to Longitudinal Research

Elisabetta Ruspini

London and New York

First published 2002
by Routledge
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by Routledge
29 West 35th Street, New York, NY 10001

Routledge is an imprint of the Taylor & Francis Group

This edition published in the Taylor & Francis e-Library, 2004.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
Ruspini, Elisabetta.
Introduction to longitudinal research / Elisabetta Ruspini
p. cm. – (Social research today)
Includes bibliographical references and index.
1. Social sciences–Research–Methodology. 2. Social sciences–Longitudinal studies.
I. Title. II. Social research today (Routledge (Firm))

H62 .R84 2002
300´.7´2–dc21 2001048954

ISBN 0-203-16722-8 Master e-book ISBN

ISBN 0-203-26205-0 (Adobe eReader Format)
ISBN 0–415–26007–8 (hbk)
ISBN 0–415–26008–6 (pbk)

Contents

Illustrations

Figures

Tables

Examples

Abbreviations

ADEPS	Analyse Dynamique des Effets des Politiques Sociales
BBS	British Births Survey
BCS70	1970 British Cohort Study
BHPS	British Household Panel Study
BSA	British Social Attitudes Survey
CAPI	Computer Assisted Personal Interviewing
CATI	Computer Assisted Telephone Interviewing
CHER	Consortium of Household Panels for European Socio-Economic Research
CHES	Child Health and Education Study
CLS	Centre for Longitudinal Studies
CNEF	Cross National Equivalent Files
CSP	Centre for Social Policy
DIW	German Institute of Economic Research
EB	Eurobarometer
ECASS	European Centre for the Analysis in the Social Sciences
ECHP	European Community Household Panel
ECPF	Encuesta Continua de Presupuestos Familiares or Household Budget Continuous Survey (see HBCS)
EDP	Echantillon Démographique Permanent
EGLHS	East German Life History Study
EHA	event history analysis
EPAG	European Panel Analysis Group
ESEML	Enquête Socio-Economique auprès des Ménages Lorrains/ Panel des Ménages Lorrains
ESSPROSS	European System of Integrated Social Protection Statistics (Eurostat)
FES	Family Expenditure Survey
GHS	General Household Survey
GHQ	General Health Questionnaire

GLHS	German Life History Study
GSOEP	German Socio-Economic Panel
HBCS	Household Budget Continuous Survey (see ECPF)
HHP	Hungarian Household Panel Survey
HHS	Department of Health and Social Sciences
HPS	Household Panel Study/Survey(s)
HRP	household reference person
HUS	Household Market and Non-market Activities
ICPSR	Inter-University Consortium for Political and Social Research
IDA	Integrated Database for Labour Market Research
ILFI	Indagine Longitudinale sulle Famiglie Italiane (Longitudinal Study of Italian Families)
INE	Instituto Nacional de Estadistica
INSEE	Direction Régionale en Lorraine de l'Institut National de la Statistique et des Études Économiques
ISER	Institute for Social and Economic Research
ISTAT	Isituto nazionale de statistica (Italian National Institute of Statistics)
LFS	Labour Force Survey
LHC	life history calendar
LII	Living in Ireland Panel Survey
LINDA	Longitudinal Individual Data for Sweden
LIS	Luxembourg Income Study
LNU	Swedish Level of Living Surveys
LS	ONS Longitudinal Study
MISEP/ ERSEP	Mutual Information System on Employment Policies
MISSOC	Mutual Information System on Social Security
MLE	maximum likelihood estimation
NCDS	National Child Development Study
NDU	National Data Collection Unit
NLS	National Longitudinal Surveys
NLSY79	National Longitudinal Survey of Youth 1979
NSI	National Statistical Institute
OECD	Organisation for Economic Co-operation and Development (based in Paris)
OLS	ordinary least squares estimation
ONS	Office for National Statistics
OPCS	Office for Population Censuses and Surveys
OR	odds ratios
OSA	Organisatie voor Strategisch Arbeidsmarktonderzoek

OSM	Original Sample Members
PACO	Panel Comparability Project
PAF	Postcode Address File
PHP	The Polish Household Panel
PSBH	Panel Study of Belgian Households
PSELL	Panel Socio Économique 'Liewen zu Lëtzebuerg/Vivre à Luxembourg'
PSID	Panel Study of Income Dynamics
QLFS	Quarterly Labour Force Survey
RLMS	Russian Longitudinal Monitoring Survey
SAS	Statistical Analysis System
SEM	Structural equation models
SEP	Belgian Socio-Economic Panel
SHIW	Bank of Italy Survey of Household Income and Wealth
SHP	The Swiss Household Panel (Vivre en Suisse – Leben in der Schweiz)
SIPP	Survey of Income and Program Participation
SIR	Scientific Information Retrieval
SLID	Survey of Labour and Income Dynamics
SOFI	Swedish Institute for Social Research
SPSS	Statistical Package for the Social Sciences
SSD	Swedish Social Science Data Research
SSI	Supplementary Security Income
STATA	software for statistical analysis
SWIP	Swedish Income Panel
TLS	Turin Longitudinal Study
TMR	Training and Mobility of Researchers
TSM	Temporary Sample Members
UDB	Longitudinal Users' Database
WES	UK 1980 Women and Employment Survey
WGLHS	West German Life History Study

Acknowledgements

This book is a synthesis of the knowledge and information acquired in the field of research and longitudinal analysis while working for my doctorate.

My interest in longitudinal research, in its forms, its aims and its potential, arose out of research done for a PhD thesis in Sociology and Social Research (University of Trento). The thesis was based on data from two Household Panel Surveys (German Socio-Economic Panel and British Household Panel Study), and focused on the gender dimension of poverty in both Germany and Great Britain. Methodological reflections on longitudinal data were further developed during post-doctorate research carried out at the University of Essex – European Centre for Analysis in the Social Sciences (ECASS) programme and at the Universities of Mannheim and Goteborg as part of the Training and Mobility of Researchers (TMR) programme set up by the European Community within the 'Family and the Welfare State in Europe' TMR research programme for young researchers. The aim of the research project carried out – 'Family, Gender and Dynamics of Poverty in Different Welfare States' – was to focus on the circumstances that explain the dynamics of lone mothers' poverty in five different European settings: Belgium, (West) Germany, Great Britain, Italy and Sweden using household panel data. My interest in longitudinal research increased during a period spent as an international visiting fellow in Social Research Methods at the University of Surrey, Department of Sociology, School of Human Sciences.

I must acknowledge the encouragement and help provided by Alberto Marradi, Martin Bulmer and Angela Dale whose comments and suggestions helped to review the first version of the book. A special thank you to Hilary Creek for her precious language revision.

This book is dedicated to my parents and to my husband, Alessandro.

Elisabetta Ruspini
Milan
October 2001

Introduction

Panta rei (everything flows)
Heraclitus

This book aims to discuss longitudinal research in its guise as a necessary tool for the study of social change, one which is particularly useful at times like today, when theoretical and methodological standardisation and stylisation are little suited to interpreting an ever more unstable, dynamic and heterogeneous society.

The problematics of change have always occupied a central position in sociological thought. The modern Social Sciences have emerged as a response to an era of very rapid, all-embracing social changes – namely the development of capitalism that destroyed the older forms of social organisation, that is, of the feudal system – and to the consequent need for greater understanding of social, economic and political processes. Indeed, one begins to study society only when it can no longer be taken for granted (Jedlowski, 1998).

Social change plays a central role in classical sociological thought. August Comte considered historical comparison to be the tool on which sociological research was based. Sociology is nothing if it is not guided by knowledge of historical evolution: 'historical comparison of the diverse consecutive states of humanity is not only the main scientific insight of the new political philosophy … it also directly forms the basis of the science, of what it can offer as being most typical' (Comte, 1842: 268). The notion of differentiation (or specialisation) was central in the work of Herbert Spencer, Emile Durkheim and Talcott Parsons. Marx described the dynamics of the capitalist system: capitalist development is achieved through expropriation of surplus value, or profit, by the capitalist, from the workers. Indeed, Marx posited contradictions and conflicts as arising from the differentiation of economic and social positions in economic systems. Max Weber established the dynamic power of culture, particularly religion, in social change (Smelser, 1981;

Haferkamp and Smelser, 1992). Furthermore, Abrams (1982) argued that sociological explanations must always be of an historical nature, because social reality is historical reality, a reality in time; while, according to Wright Mills (1959) 'social science deals with the problems of biography, history and of the way they affect the body of social structures'.

Thus, change is a prime feature of a social reality that any social-scientific theory must, sooner or later, address. However, even though the analysis of social change represents the touchstone of sociology – and though the subject studied in sociology is continuously undergoing transformation – the study of social change clearly has, so far, not been developed to its fullest extent (Wiswede and Kutsch, 1978). This lack could depend on the combination of two elements: one theoretical and one methodological. First, the apparent difficulty of reconciling theories about social change – developed at the macro-sociological level – with the changing life-course patterns of individuals and with opportunities for analysis offered by empirical research (from the use of documents and empirical analysis of life histories, to panel studies); and second, the lack of longitudinal information about the social-demographic characteristics of both individuals and households (prospective longitudinal data on households became available in the 1970s in the United States and only in the 1980s in Europe)[1] and of techniques designed to manipulate the longitudinal dimension.

Many European longitudinal studies were set up in the 1980s, when both state and private agencies began to provide considerably more funding for the collection of nationally representative longitudinal datasets. Indeed, since the 1970s, all advanced industrial societies had been undergoing a period of profound socioeconomic change: the differentiation and instability of family models and the consequent erosion of the protective role of the nuclear family; the growing importance of the service sector; the decline of secure employment both in large manufacturing industry and in the tertiary sector. Alongside these there had been increases both in the number of people experiencing either prolonged periods of unemployment or definitive ejection from employment – particularly among some social groups such as women or youth – and unstable, atypical, temporary, very low paid jobs. There are many approaches to describing the current changes in the world, such as the transformation to a knowledge-based society, globalisation, post-industrialisation, post-Fordism, late-, reflexive- or post-modernity (see, among others, Bell, 1973; Touraine, 1974; Giddens, 1990; Bauman, 1992; Beck, 1992).

These changes created serious difficulties for existing social support systems which had originally been developed on the basis of very different life styles, of different forms of family organisation and on a marked diversity, and division, between male and female gender roles. These social support systems had, in fact, been structured on the basis of the assumption that there would

continue to be more or less steady, linear, economic growth which would be able to absorb the new labour power and to maintain full employment. These systems had, to a large extent, been designed to deal with only brief periods of unemployment and relied on the assumption that the labour market would, sooner or later, always offer job opportunities to workers who had been made redundant.

In this fluid and complex situation, the risk of both social precariousness and of poverty, both increased and diversified. Consequently, it soon became clear that new methods of inquiry, new tools, were required to help understand not only such fast social change, but also the dynamic nature of people's lives and the radical shifts taking place in the role of social institutions. In particular, interest in the link that unites the descent into poverty with macro and micro social processes, and the consequent need to follow up the biographies of subjects in difficulty – with the aim of understanding the genesis and the dynamics of such processes and, thus, forestalling the risk of deprivation – have encouraged a move towards more systematic data-gathering and the development of techniques that permit dynamic interpretation of the processes of social exclusion and poverty. To offer just one example, the main aims of Household Panel Studies/Surveys (HPSs) (here, household means a cohabiting group/unit) are to analyse fluctuations in income and to describe and explain changes in the economic situation of the subjects studied.

In Europe, it is now becoming easier to access prospective and retrospective longitudinal data. This will make it possible to develop an analytical prospective of life-courses and constitutes one of the most important developments of official statistics in the last two decades (Ghellini and Trivellato, 1996). But there is still a large gap between this increasing availability and everyday research practices which, today, are still largely restricted to cross-sectional type analyses. More specifically, in Europe, longitudinal studies are not necessarily being developed at the same pace: the very high cost of such studies has made it difficult to set them up in countries with more economic problems. It is no coincidence that the countries where such research has been established longest are all in Northern Europe (Germany, Sweden, Holland, Belgium, Great Britain). Indeed, the oldest and most important examples of HPSs are: the Panel Study of Income Dynamics (PSID) in the US, the German Socio-Economic Panel (GSOEP) and the British Household Panel Study (BHPS) in the UK. The gap in the availability of longitudinal datasets between Northern and Southern Europe is thus evident. In the countries of Southern Europe, where there is no tradition of longitudinal research, panel data are markedly slow to become available. For example, in Italy the availability of longitudinal datasets is very limited: the Italian Official Statistics, today, produce very little longitudinal data.[2]

A number of factors can explain the relative lack of available panel data and the small amount being produced: factors such as the high costs of gathering such data, the complexity of the data-gathering process, the complexity of the data gathered, the fact that much of this information is confidential, which often inhibits the public distribution of such data.[3] Consequently, in many areas (e.g. those in Southern Europe) longitudinal research is little used in social research notwithstanding the pressing need to produce such data and to make them available. Longitudinal data are essential if a researcher wishes to measure social change and evolution through history. Moreover, as Rajulton and Ravanera (2000) argued, in spite of the general acceptance of the usefulness of longitudinal data, many researchers are still not ready to adopt suitable techniques of analysing such data. This situation cannot be rectified unless we find a way to disseminate techniques of analysis to would-be users of longitudinal data.

Longitudinal analysis is, simultaneously, a necessity, a luxury and a riddle for the social sciences (Mingione, 1999). It is a necessity because one presumes that the actor's experience, including the length of the period and the precise historical point when the experience took place, would have a determining influence on his/her behaviour. It is a 'luxury' in two senses: longitudinal data are very expensive to gather and the human costs involved in interpreting the results are high. Lastly, it is a riddle because of the sheer complexity of cross-sectional information about historical events: longitudinal studies multiply information because the variables take on different meanings at different historical moments. Longitudinal research is also a challenge. As Leisering and Walker stated:

> approaches to 'thinking dynamically' have triggered the beginnings of an intellectual revolution, one that blends insights from across the social sciences, merges quantitative and qualitative methodologies, combines macro and micro views of society and exploits the power of international comparisons.
>
> (1998a: xiv)

Hence it is becoming important to construct, and to encourage, wider use of longitudinal methodology – which places a high priority on longitudinal research and could help in designing and setting up research activities – and, simultaneously fully exploiting, and making more available, the few existing examples of extant longitudinal surveys. To do this there must be an exchange of information between those who have already worked and reasoned 'longitudinally', those who would like to do so but are not sure how and those who are wary of the consequences of approaching and dealing with dynamic data.

This book is an invitation to use longitudinal research as a basis for a better understanding of the processes of social change. The idea is to persuade the reader of the value and advantages of carrying out research 'longitudinally' and to encourage both students and researchers to create their own longitudinal research projects.

The aim is to offer guidelines to anyone who needs to carry out research based on longitudinal data (on trends or repeated cross-sectional, panel and duration data/event history data) and to promote a better understanding of the type of information that longitudinal data provide and of the techniques needed to analyse such data. Both the clear advantages and the problems of using dynamic data are highlighted, as are also the potential benefits they offer. The availability of longitudinal datasets in Europe is also discussed. Moreover, some useful paradigms and initial steps to be taken when undertaking longitudinal analysis are presented. Lastly, some of the inherent implications of integrating empirical research on social change conducted at the micro level and theories on macro-type change are explored as is the dialogue between them (from *macro* to *micro* social change and *vice versa*).

The volume is divided into two parts: longitudinal research and longitudinal analysis. The dynamic approach is both a paradigm and a method. It offers both the theoretical framework needed to explore the dynamic character of society and also provides a method needed to 'capture' these dynamics. Chapter 1 analyses the concept of longitudinal research. Chapters 2 and 3 present the features of existing dynamic files and raise the crucial problems of the availability of such data and of comparability within longitudinal research. Chapter 4 examines the problems that may emerge when dealing with this type of data. Lastly, Chapter 5 offers a very clear, user-friendly overview of the analytical techniques usually employed with longitudinal data: here, the level of statistical complexity is kept to a minimum. A presentation of the most salient features of existing longitudinal files concludes the work. The list of abbreviations will help the reader to sort out what the different longitudinal studies and their acronyms are. Throughout the text the reader will find examples which offer further in-depth explanations of the concepts, methods and techniques used and described in the book.

Notes

1 Prospective longitudinal studies are usually based on a probability sample of individuals/families and carried out by means of repeated interviews at fixed intervals.

2 Only three microlevel longitudinal studies have ever been carried out on a nationwide sample in Italy. In chronological order, these are:

- The Bank of Italy Survey of Household Income and Wealth (SHIW), which started in 1965 and continued, unchanged, until 1987, and is based on data gathered independently

at different points in time. In order to facilitate analysis of the way in which any of the phenomena identified were evolving, a new technique was incorporated into the 1989 survey, one which would be able to take into account the fact that some of the individuals in this current sample had already been interviewed in previous surveys (Brandolini and Cannari, 1994).

- The European Community Household Panel (ECHP), was set up in 1994 and is based on a Europe-wide probability sample of 60,819 households drawn from member states in a proportion which reflected their population.
- The Longitudinal Study of Italian Families (Indagine Longitudinale sulle Famiglie Italiane or ILFI), is a prospective panel study. The first retrospective survey was carried out in 1997 by the University of Trento, the Istituto Trentino di Cultura (the Trento Cultural Institute) and ISTAT (Italian National Institute of Statistics) and based on a nationwide sample of 4,714 families (10,423 individuals). The second, prospective wave ended in 1999, while the third wave (2001) is, currently, being launched.

3 For example, even today the rules for accessing Europanel data are quite restrictive (see Appendix 2 for details). These problems can be resolved, while still respecting privacy, through the excellent work of collecting and distributing files of data held in the diverse archives to be found in many European countries: Austria (WISDOM), Belgium (BASS), Denmark (DDA), France (BDSP), Germany (ZA), Hungary (TÁRKI), Italy (ADPSS), The Netherlands (STAR), Norway (NSD), Sweden (SSD), Switzerland (SIDOS) and the United Kingdom (UK-DA). These archives have been organised into a consortium (Council of European Social Science Data Archives), which aims to act as an international network for promoting and facilitating the exchange of data required for research. Web site: http://www.nsd.uib.no/cessda/europe.html.

Part I

Longitudinal research

1 What is longitudinal research?

The term 'longitudinal' will be used here to describe what can be defined as the minimum common denominator of a family of those methods which tell us about change at the individual micro level (Zazzo, 1967; Menard, 1991). This family is the opposite of that described by the term 'cross-sectional research'.

'Longitudinal' is a rather imprecise term. Longitudinal data can be defined as data gathered during the observation of subjects on a number of variables over time. This definition implies the notion of repeated measurements (van der Kamp and Bijleveld, 1998). Basically, longitudinal data present information about what happened to a set of units (people, households, firms, etc.) across time. The participants in a typical longitudinal study are asked to provide information about their behaviour and attitudes regarding the issues of interest on a number of separate occasions in time (called the 'waves' of the study) (Taris, 2000). In contrast, cross-sectional data refer to the circumstances of respondents at one particular point in time (I shall expand on these points later). Thus, the term 'longitudinal' refers to a particular type of relationship between phenomena: the type which evolves over the course of time and is termed *diachronic*, the opposite of *synchronic*.

There are many different methods that can be used to collect longitudinal data, which means there are also many different types of research (Buck *et al.*, 1994; Davies and Dale, 1994; Bijleveld *et al.*, 1998; Ruspini, 1999, 2000a; Taris, 2000).

The most commonly used longitudinal designs are:

- repeated cross-sectional studies (trend), carried out regularly, each time using a largely different sample or a completely new sample;
- prospective longitudinal studies (panel), that repeatedly interview the same subjects over a period of time;
- retrospective longitudinal studies (event history or duration data) in which interviewees are asked to remember, and reconstruct, events and aspects of their own life-courses.

Of these three, prospective studies are considered the most 'truly longitudinal' (consequently preferable when analysing microsocial change), because they, periodically, gather information about the same individuals (Janson, 1990; Magnusson *et al.*, 1991), who are asked the same sequence of questions at regular intervals. In particular, prospective longitudinal surveys provide the most reliable data on change in knowledge or attitudes, because longitudinal measures are collected while the subjective states actually exist. Indeed, some consider retrospective surveys to be 'quasi-longitudinal', both because they offer only an incomplete contribution to the study of causal processes and, above all, because of distortions due to inaccuracies in memories (Hakim, 1987: 97; Draper and Marcos, 1990; Dex, 1991; Taris, 2000).

Each longitudinal design will be examined separately here (see Chapter 2 for details).

A *cross-sectional survey* studies a cross-section of the population at a specific moment or point in time. Here, the term 'cross-section' indicates a wide sample of people of different ages, education, religion and so on. Repeated cross-sectional studies, such as the General Household Survey or the Family Expenditure Survey in Great Britain, the European Community Eurobarometer Surveys, the Italian National Institute of Statistics (ISTAT) Multipurpose Survey of Italian Families (Indagine Multiscopo sulle famiglie italiane) and the Bank of Italy Survey of Household Income and Wealth, can help in the study of social change. However, because these surveys are not based on the same sample, they only offer a means for analysing *net changes* at the aggregate level – the *net effect* of all the changes – (Firebaugh, 1997): e.g. a comparison between the incidence of poverty and the characteristics of the population below the poverty line at time t and at time t–1 or between the pool of employed and unemployed in two different years. Thus, cross-sections can tell us about populations either at one or at a series of points in time.

Longitudinal data tell us about change at the individual or micro level providing estimates of both net and gross change – that is, the analysis of flows between states – and other components of individual change (i.e. to disaggregate net change) (Rose, 2000: 27). *Prospective longitudinal studies*, especially Household Panel Studies (HPS), follow individuals and families over time by, periodically, re-interviewing the same subjects and providing multiple observations on each individual/household in the sample. Such studies involve not only a random sample of households, but also all those members and subsequent co-residents, partners and descendants who are repeatedly re-interviewed. Thus, these studies accumulate records of employment, income, family status and attitudes over extended periods. This makes it possible to study change at the individual, i.e. the *micro*, level (Hakim 1987; Rose and Sullivan, 1996; Gershuny, 1998, 2000), that is, to analyse changes within the institutional, cultural and social environments that surround the individual and shape the course of his/her life. Thus, they

offer a basis for further study of the dynamics of social phenomena – an advantage Paul F. Lazarsfeld must have recognised when, in the late 1930s, he was the first to use longitudinal data when analysing the relation between radio advertisement and product sales/changes in public opinion. He suggested repeatedly interviewing the same respondents would clarify whether the radio advertisement was the cause or the effect of buying the product (Lazarsfeld, 1940). Thus, for Lazarsfeld the panel technique seemed to be one of the most promising for the future of a fuller understanding of human behaviour (Lazarsfeld, 1948).

Among the areas of panel research which have been identified as being of particular concern to policy-makers are the following (Rose, 2000):

- dynamic analyses of labour income (Joshi and Davies, 2000);
- analysis of career trajectories (Scherer 2000; Gallie and Paugam, 2000);
- poverty and income dynamics (Walker and Ashworth, 1994; Ashworth *et al.*, 2000; Jarvis and Jenkins, 2000; Muffels 2000);
- the gender dimension of poverty (Ruspini, 2000b);
- child poverty, child achievement and parenting (Ashworth *et al.*, 1992a, 1992b; Hill and Jenkins, 1999);
- well-being of the elderly (Coe, 1988; Burkhauser and Duncan, 1988, 1991; Bound *et al.*, 1991; Lillard and Waite, 1995);
- social exclusion (Walker, 1995);
- analysis of welfare use (Walker and Ashworth, 1994);
- analysis of the achievements and failures of welfare states (Goodin *et al.*, 1999);
- household change: household formation and dissolution (Blossfeld, 1995; Jarvis and Jenkins, 1998; Ermisch, 2000);
- dynamic issues of disability[1] (Adler, 1992; Eustis *et al.*, 1995);
- transitions, e.g. into/out of the labour force; from youth to adulthood.

Event history or *duration data* offers a record of the events that have punctuated the life-course of a group of subjects. These concepts need to be clarified. *Life-course* is used to refer to the history of each family or individual and to the way this history evolves and changes over time (Saraceno, 1986). The life-course is determined by interdependent *trajectories* and *transitions* that subjects (individual or collective – woman, man, couple, firm) undergo during the course of their existence. Trajectories refer to the path taken, as time goes on, within a specific, relatively long-term experience or position – the family, work, etc. – one which often may continue for a large part of the individual's lifespan. Transitions are fluctuations/changes within a trajectory: in other words, trajectories are characterised by the transitions, or changes, of social, economic and demographic interests which evolve in response to specific *events* (Elder, 1985). In this instance, 'event' is taken to mean a change, or a transition, from one discrete state to another, a

passage which takes place at a specific point in time and which constitutes a radical departure from what came before the 'catalysing' event: e.g. marriage, the birth of a child, starting work, divorce, etc. (Allison, 1984). Thus, an event can be defined as a change that gives an individual new status, which differs from the previous status the individual had before the change took place.

This definition of an event enables us to visualise events as transitions between states (Rajulton, 1999). The most important transitions (e.g. the transition to adulthood) usually introduce a multiplicity of changes into individuals' lives (Billari, 1998). However, apparently similar transitions may assume a different significance depending on the point at which they take place within a particular trajectory: going to university straight from school or after taking a few years out to work; having a child at 20 or at 40; being made redundant when a young adult and losing a job when middle-aged with adolescent children to support (Olagnero and Saraceno, 1993). Thus, life-course dynamics arise from the interplay of trajectories and transitions, an interdependence played out over time and in relation to others (Elder, 1985).

Duration data are usually gathered using retrospective cross-sectional studies in which respondents are asked to remember events and aspects of their own life-courses. Typically, this is done domain by domain, beginning with the current situation and taking respondents backwards in time. In panel surveys, data may be collected at the first wave either retrospectively for a fixed initial reference period or as far back as a specific event, such as marriage or first employment (Skinner, 2000). While this design is both simple and cheap, these data are typically more complicated than those obtained with trend or panel techniques because detailed information is given for each episode – that is, a time span a unit of analysis (e.g. a woman/man) spends in a specific state – details about the duration and frequency of the event and about any other aspects which show marked diachronic variation. However, retrospective surveys do have clear limitations, both in the necessarily simplified form in which they are forced to reconstruct experiences and, above all, because memory often distorts reality when trying to recall past events (Dex, 1991). Hence, retrospective surveys are usually limited to significant but infrequent life events such as births, marriages, divorces and job changes (Rose, 2000: 12).

Research is rarely based on one investigative method alone; indeed, longitudinal research is commonly based on a mix of methods.[2] Some examples of longitudinal mixed designs are:

1 Repeated cross-sectional studies one part of which are done in the form of panel studies. For example, the British Social Attitudes Survey (BSA) or the Bank of Italy Survey of Household Income and Wealth (SHIW)

are repeated regularly on a largely different sample but with a small part as a panel study (Jowell *et al.*, 1992).

Example 1.1 Examples of repeated cross-sectional studies with a small panel section

The BSA is an annual survey that measures changes in social attitudes, with particular reference to influence upon the ways that people vote, which has been charting changing values in Britain since 1983. Core-funded by the Sainsbury Family Charitable Trust, its findings are based on hour-long interviews with a sample of 3,600 people. The survey is designed to yield a representative sample of adults aged 18 and over in England, Scotland and Wales. Since 1993 the sampling frame for the survey has been the Postcode Address File (PAF), a list of addresses (or postal delivery points) compiled by the Post Office. The sample is confined to those living in private households. People living in institutions are excluded, as are households whose addresses are not on the PAF. In most years three versions of the BSA questionnaire are fielded. Each 'module' of questions is asked either of the full sample (around 3,600 respondents) or of a random two-thirds or one-third of the sample. Two of the main purposes of the BSA series is to allow monitoring of patterns of continuity and change, and the examination of the relative rates at which attitudes change over time with respect to social issues. The subjects covered by the surveys are wide-ranging, but include housing and home ownership, work and unemployment, health and social care, education, business and industry, social security and dependency, tax and spending, the welfare state, transport, environment and the countryside, constitutional reform, law and order, civil liberties, moral issues and sexual mores, racism and sexism, social inequality, religion, politics and governance.
Web sites: http://qb.soc.surrey.ac.uk/surveys/bsa/bsaintro.htm
http://www.data-archive.ac.uk/findingData/bsaAbstract.asp

The SHIW was launched in 1965. Twenty-three further surveys have been conducted since then, yearly until 1987 (except for 1985) and every two years thereafter. The aim of the survey is to gather information about the economic behaviour of Italian families at the microeconomic level. Data on family income, saving, expenditure, consumer durables and real wealth have been collected since 1966, while the acquisition of details concerning total consumption expenditure started in 1980. The basic survey unit is the household, which is defined in terms of family relationships, that is, as a group of individuals linked by ties of blood, marriage or affection, sharing the same dwelling and pooling all or part of their incomes. Persons living in nursing homes for the aged or ill, in prisons or military installations are not included. The survey has a panel section, corresponding to: 15.0 per cent of the households between 1987 and 1989; 26.7 per cent between

1989 and 1991; 42.9 per cent between 1991 and 1993; 44.8 per cent between 1993 and 1995, 37.3 per cent between 1995 and 1998 and 48.4 per cent between 1998 and 2000 (Brandolini and Cannari, 1994; D'Alessio and Faiella, 2002).
Web site: http://www.bancaditalia.it

---

2 Prospective studies that gather information systematically through the use of calendars and/or suitable batteries of questions which aim to retrospectively investigate the life of the interviewee but not necessarily enquire about the same subject each time. One typical example are Household Panel Studies (HPSs), the most important of these being the Panel Study of Income Dynamics (PSID) in the United States, the German Socio-Economic Panel (GSOEP) in Germany and the British Household Panel Study (BHPS) in Britain.

Example 1.2 Household Panel Studies

The PSID is the longest running household panel today. It is a prospective longitudinal study, set up in 1968 in the Survey Research Center – Institute for Social Research (University of Michigan) and based on a proportional sample of the resident population of the United States (men, women and children) and their families. Since 1985, the PSID has also been collecting detailed, retrospective data on the histories, both family and matrimonial, of the subjects in the sample (the Demographic History Files).
Web site: http://www.isr.umich.edu/src/psid/

The GSOEP is a representative longitudinal study of private households in the Federal Republic of Germany. It has been modelled on the PSID. Its first wave went into the field in 1984, with a sample of 5,921 households and 12,245 individuals. The same private households, persons and families have been surveyed annually since 1984. The GSOEP has been developed and is carried out by the Project Group 'Socio-Economic Panel' at the German Institute for Economic Research (DIW), Berlin. In co-operation with the DIW, the Centre for Policy Research at Syracuse University has prepared an English language public-use version of the GSOEP for use by the international research community. The public-use version of the GSOEP is offered to researchers throughout the world for use when studying the socio-economic characteristics of persons living in Germany (Butrica, 1996a). In order to reduce the risk of identifying individuals or households, this file does not include detailed information on nationality or region and represents a 95 per cent random sample of the original data. GSOEP data cover a wide range of subjects including: household composition; occupational and family biographies; employment and professional mobility; earnings; health; personal satisfaction as well as subjects covered in topical

modules of the survey. Topical modules add questions on a variety of topics not covered in the core section: social security; education and training; allocation of time; family and social services. Moreover, two calendars are included in the core questionnaires: these calendars record monthly retrospective information on labour force participation and income.
Web sites: http://www.diw.de/english/sop/index.html
http://www.diw.de/english/sop/uebersicht/

The first three waves of the BHPS gathered information about employment histories, both family and matrimonial histories, and on individual demographic behaviour. The research group which conducts the BHPS has recently created a file, the BHPS Work-Life History Project, which puts together prospective and retrospective data (gathered during the second wave) concerning the employment conditions and the work/employment histories of those interviewed. This has made it possible to trace and reconstruct the occupational biographies of interviewees from the moment they entered the labour market up to the time of the most recent wave. More precisely, the Work-Life History Project is based on all sources of employment status and occupational information in the BHPS. These files combine information from:

- the inter-wave job history, all waves;
- the main file for current individual status, all waves;
- retrospective occupational history, wave 3;
- retrospective employment-status history, wave 2.

Web site: ://www.data-archive.ac.uk/doc/3954%5Cmrdoc%5Cpdf%5Cnewman.pdf

3 Cohort studies that are also prospective and/or retrospective (two British examples of this being the National Child Development Study and the Birth Cohort Study). Typically, in a cohort study one or more generations are followed over time, that is, over their life-course. A cohort has been defined as 'the aggregate of individuals who experienced the same life event within the same time interval' (Ryder, 1965: 845) birth, marriage, moment of entry in the labour market, moment of diagnosis of a particular disease, etc. One particularly important type of cohort is the 'birth cohort', that is the set of people who were born in the same year. Thus, cohort studies may begin at birth, but may also begin at a much later age (Bynner, 1993; Davies and Dale, 1994; Taris, 2000) (see Chapter 2 for details).

Example 1.3 Examples of cohort studies

The National Child Development Study (NCDS) is a multi-disciplinary longitudinal study which takes as its subjects all those living in Great Britain who were born

between 3 and 9 March 1958 (about 17,000 individuals). To date, there have been six attempts to trace all the members of the birth cohort in order to monitor their physical, educational and social development: one in 1965, when they were aged 7, one in 1969, when they were aged 11, one in 1974, when they were aged 16, one in 1981, when they were aged 23 and then in 1991, when they were aged 33. A sixth sweep was conducted in 1999, and will soon be available for analysis. In addition, in 1978, contact was made with the schools and colleges they attended. Information was obtained from the mother and medical records from the midwife, along with various information acquired throughout the individual's life. The NCDS is used for a wide range of research, including medical/health research. The data cover a long-term period and include a wide range of questions, plus physical measurements such as weight and height. The aim of the study is to improve understanding of the factors affecting human development over the whole lifespan.
Web site: http://www.cls.ioe.ac.uk/Ncds/nintro.htm

One example of a multi-purpose longitudinal study starting later than birth is the Swedish Malmö Study which started in 1938 with a sample of 1,500 children in the third grade of school (average age 10). The sample has been followed up through six surveys into adulthood with over 1,000 still participating. The scope of data collection is much the same as in the British birth cohort studies, but with more psychological emphasis: psychological well-being, health and social network were covered in childhood and family formation, occupation, income and health were covered through adulthood (Furu, 1995; Furu and Hellström, 1996; Hellström 1996). Finally, an older age group, comprising a nationally representative 'panel' selected from the total Swedish population has been surveyed in the Swedish Level of Living Surveys (LNU), based in the Swedish Institute for Social Research in Stockholm University. This started in 1968 and has involved following up 9,741 cohort members, in the age band 15–75 over four sweeps (1968, 1974, 1981, 1991). Over 7,500 are still participating. The main topics covered are health status, working conditions, economic resources, housing standards, family, social integration, education and employment (Erikson and Åberg, 1987; Johansson, 1973; Tåhlin, 1990).
Web site: http://www.ssd.gu.se/kid/swe/lnu.html

To sum up, longitudinal research collects information about the temporal evolution of behaviour and ensures that the same individual will be involved each time. Where individuals are surveyed at successive time points, then it is possible to investigate how individual outcomes are related to the earlier circumstances of the same individuals. Thus, longitudinal data not only begin to unravel the nature of change at an individual level but also present opportunities to recognise explicitly that individual behaviour is characterised

by strong temporal tendencies. Longitudinal data then become essential if we are to understand these temporal tendencies in micro-level behaviour (Davies and Dale, 1994).

The development of longitudinal research: an historical overview

Longitudinal data have been being collected for a long time, but the idea of using such data for research purposes is a recent development, especially in the field of sociology.

Menard (1991) wrote that longitudinal data have, in fact, been collected at the national level for more than 300 years. The first regular, periodical collections of census data were carried out in Quebec (CDN) when it was still a French colony called New France (1665–1754). Other long-running periodical censuses which should be mentioned are: Sweden (started 1749); Norway and Denmark (1769) and the United States (1790). In Italy, the first census was launched in 1861, the year of unification.[3]

At the more 'properly' longitudinal level (in the sense of studies designed to gather data about the dynamics of individual phenomena, on a regular basis) diachronic studies are mostly to be found in the fields of medicine, psychology and anthropometry. Wall and Williams (1970) and Nesselroade and Baltes (1979) have argued that biographical data about individuals were collected for the first time in the eighteenth century. Some early examples of such research which are well worth mentioning are: that carried out by De Montbeillard – who, between 1759 and 1777, recorded the stages of growth of his own son from birth to 18 – and, much more recently, studies by Tiedmann and Shinn on the development of sensory perception during the first three years of life (Shinn, 1907).

The United States played a pioneering role in the development of longitudinal research. Indeed, the earliest attempts to gather and analyse dynamic data and, simultaneously, to use biographical data were all made in the United States where, by the late 1920s/early 1930s, many longitudinal studies on childhood were already well under way (for details see Wall and Williams, 1970; Mednick and Mednick, 1984). Many, but not all, of these studies concentrated on the evolution of children's physical characteristics (Sontag, 1971; Kessler and Greenberg, 1981). Among the 'classic' longitudinal studies of human development sponsored by the National Research Council, one should certainly mention that by Terman, begun in 1921, which aimed to study the physical, mental and personality development of gifted children (Terman, 1925, 1939; Terman and Oden, 1947, 1959), and studies of psycho-physical development which were launched at the Merrill Palmer School in Detroit (1923), the Medical School of Colorado University (1923) and the

University of Minnesota (1925). In 1928 the Berkeley Growth Study began; in 1929 the Berkeley Guidance Study, the Fels Research Institute Project, and the Harvard Growth Study were launched. The final project of this series was the Oakland Growth Study, which was initiated in 1932. All these studies were global in their approach: to give an example, the study by Terman maintained a continuing record at approximately 10-year stages on physique, health, personal and social adjustment, nature of interests and activities – with detailed educational, vocational and marital histories – of 1,528 children (857 males and 671 females) selected from the State of California. As another example, the Harvard Growth Study – entitled Longitudinal Study of Child Health and Development and conducted by the Harvard School of Public Health (1929) on a group of 309 new-born children – collected about 200,000 observations over a period of 18 years and then launched a follow-up, on themes connected to health and social relations, when the subjects were aged between 25 and 34. It involved an interdisciplinary team of medical, biological and social scientists.[4]

There are certain similarities between these pioneering studies: marked multi-disciplinarity, the use of anthropometric tools, analysis of both the individual's physical development and of the evolution of his/her personality and, lastly, the fact that they all recorded information about family situations and environments. However, most of these surveys lacked stated hypotheses, had selected a methodology without delineating specific problems (Mednick and Mednick, 1984), and were based on only a small number of subjects, not chosen on probabilistic grounds, but on the basis of criteria such as how close the subjects lived to the research centre and/or on their willingness to co-operate. For example, the Berkeley Guidance Study was based on a group of white, middle-class volunteers. Another problem for these surveys was the considerable attrition rate, which was sometimes more than 30 per cent. The attrition rate is defined as the measure of the degree of success in interviewing the same set of units over time: some people participate in the initial assessments but then drop out of the study. This poses the question of whether they are different in some important way from the ones who stayed with the study throughout (Copeland and White, 1991: 21) (see Chapter 4 at pp. 71–2 for details).

Example 1.4 The Berkeley Guidance Study and the Oakland Growth Study

During the late 1920s and early 1930s, two pioneering studies of children were launched at the Institute of Child Welfare (now Human Development) at the University of California, Berkeley. The Berkeley Guidance Study, under the direction of Jean Macfarlane, started with a sample of 248 infants who were born in Berkeley, California in 1928–29. This sample was divided into two groups; an intensively studied group which provided detailed annual information on socio-

economic conditions and family patterns, and a less intensively studied 'control' group which was matched on social and economic characteristics. Most of the children were Caucasian and Protestant, and two-thirds came from middle-class families. The basic cohort included 214 of these children and their families who participated in the study throughout the 1930s and up to the end of the Second World War. Annual data collection ended in 1946, but there were two adult follow-ups (1959–60 and 1969) in which most of the children participated.

The Oakland Growth Study, under the direction of Harold Jones and Herbert Stolz, was launched in 1931 to study the physical, intellectual and social development of boys and girls, and commenced data collection in 1932. The 167 children who were intensively studied from 1932 to 1939 were initially selected from the fifth and sixth grades (birth years, 1920–21) of five elementary schools in the north-eastern section of Oakland, California. There were five waves of data collection during their adult years, finishing in 1980–81: these follow-ups generally included interviews, health assessments, personality inventories and fact-sheet questionnaires. Elder's most famous book (1974) was based on his work with the Oakland cohort. In that book, he combined an historical, social and psychological approach to assess the influence of the economic crisis on the life-course of those 167 people born in 1920–21.

Web site: http://www.cpc.unc.edu/lifecourse/berkoak.html

Gathering repeated data periodically from, and about, the same individuals was thus common practice long before the term 'panel' began to be used by the scientific community. It was Lazarsfeld who first introduced the concept of panel when, during the 1940s, he was investigating the ways in which causal relations could be identified and the problem of ambiguity within the causation process (in this case, the relationship between radio advertising and product sales). He suggested that repeated interviews, with the same subjects but at different points in time, might be able to reveal whether listening to a particular radio advert was the cause or the effect of any subsequent purchase made of a specific product (Lazarsfeld and Fiske, 1938; Lazarsfeld, 1940, 1948, 1972).

Event history analysis (EHA), that is, the study of duration data – or rather of the time that elapsed before an event (which could be bereavement, a new marriage, the birth of a child, divorce, etc.) – was also initially developed in the United States, in the fields of biomedicine and engineering. In the first case, doctors were interested in the time that had elapsed between the administration of a drug and the death of the animal used as a guinea pig (i.e. the survival time): the event studied being the death of the animal (Gross and Clark, 1975). Such studies are termed survival analysis or lifetime analysis and they make ample use of survival or life tables[5] (which have been in use since the seventeenth century) in order to study how long the animal being treated survives (Kalbfleisch and Prentice, 1980). A similar

approach can be found in the field of engineering (Barlow and Proschan, 1975). Here the problem studied was the process of deterioration of both machines and electronic components, and the technique used was termed either reliability, or failure time, analysis (Nelson, 1982). Because it was developed in different contexts but at much the same time, the most important concept of EHA, the propensity to change state, has been given many names: transition rate, hazard rate, intensity rate, failure rate, transition intensity, risk function or mortality rate (Blossfeld and Rohwer, 1995: 28).

In the 1920s and 1930s, social conditions in the United States also encouraged the development of biographical research.[6] In this period, particularly important research on biographies was carried out by the Sociology Department of the University of Chicago which, primarily, launched these surveys to study urban marginalisation.

Scholars of the Chicago School developed an approach that offered an alternative to the traditional way of studying hardship when they sought to describe scenes of deprivation, not only from the point of view of economic hardship but also from the prospective of other different processes such as social exclusion (Micheli and Laffi, 1995).[7] The innovative aspect of this approach lies in the accurate analyses carried out on deviant or marginalised figures, their group dynamics, their relationships with the community and with their immediate environment (Anderson, 1923). In this way each individual's biography was built up on the basis of a wide range of sources of information: reconstructions, accounts given by the subjects themselves, personal documents and participant observation.[8]

As Olagnero and Saraceno (1993: 25–7) explained, the aim was to study new social actors, diverse or deviant, through the information gleaned from their individual biographies, to study those actors who had, until then, been left on the sidelines by the functionalist approach: thus there was a change of direction, from a sociology based on structure and functions to a sociology of relations. In this way, increased attention could be paid to the dynamics and the processes of social life and an attempt was made to study life histories over time within the context of a constantly changing environment.

However, empirical research was slow to take its cue from these important reflections, reflections which already offered many of the elements necessary for developing adequate analyses of life histories. Until the 1980s, for example, the study of deprivation continued to be dominated by a static vision of poverty which was defined as a condition in itself linked to a specific moment in time. Only recently has the debate on poverty begun to focus on two complementary themes: on one hand, the relation between poverty and the way in which welfare systems are organised and, on the other, the dynamic aspects of poverty, i.e. on reconstructing the path that leads to poverty (Ruspini, 2000b).

The diachronic nature of phenomena such as deprivation or dependence on welfare began to be recognised thanks to the development of HPSs. The results obtained from such studies encouraged a radical change in the way in which the phenomenon was perceived. However, these results only really began to become available in the late 1960s in the United States and in the late 1980s in Europe. Thus, in this case too, longitudinal research on prospective data was first developed in the United States where it was encouraged by the strongly pragmatic orientation of North American Sociology, noted for its propensity to concentrate on analysing and solving social problems. Indeed, the first household panel in history was launched in the United States in 1968: the legendary PSID, which provided the inspiration for all subsequent HPSs. One of the motivations for this project was the assumption that poverty was self-perpetuating. The panel design offered a way to determine whether such views corresponded with reality (Elder, 1985). Contrary to prevailing beliefs at the time, only a very small fraction of sample members who actually experienced poverty did so beyond a year or more. The same was true for welfare dependency: welfare recipients remained on the welfare rolls for relatively short periods of time (Coe, Duncan and Hill, 1982). In Europe the earliest longitudinal studies of the family, the GSOEP (Germany) and the BHPS (Britain) were directly inspired by the US example.

Example 1.5 The PSID

Starting with a national sample of approximately 4,800 US households in 1968, the PSID has re-interviewed individuals from those households every year since that time, whether or not they have been living in the same dwelling or with the same people. Adults have been followed as they have grown older, and children have been observed as they advance from childhood to adulthood, forming households of their own. In 1990, a representative national sample of 2,043 Latino households, differentially sampled to provide adequate numbers of Puerto Rican, Mexican-American, and Cuban-Americans, were added to the PSID database. The PSID provides a wide variety of information both about families and their individual members, with some information about the areas where they live. The central focus of the data is economic and demographic, with substantial details on income sources and amounts, employment, family composition changes and residential location.

Web sites: http://www.isr.umich.edu/src/psid/
http://www.isr.umich.edu/src/psid/overview.html

Cohort studies were already under way some time before the HPS technique was developed. Many had begun to be set up much earlier, from the 1950s on, especially in the Anglo-Saxon world: studies such as the already

cited British NCDS, launched in 1958, the National Longitudinal Surveys (NLS) (1966) which were set up in the United States, and the British Cohort Study (BCS70), which started in 1970. All these studies aimed to study one or more cohorts of individuals over time in order to monitor both their individual, and social, growth and development.

Example 1.6 The NLS and the BCS70

The NLS sponsored and directed by the Bureau of Labor Statistics, US Department of Labor, gather detailed information about the labour market experiences and other aspects of the lives of six cohorts of women and men. The surveys include data about a wide range of events such as schooling and schooling to career transitions, marriage and fertility, training investments, child-care usage and drug and alcohol use. Thus, each survey allows for analysis of an extensive variety of topics such as the transition from school to work, job mobility, youth unemployment, educational attainment and the returns to education, welfare recipiency, the impact of training and retirement decisions.

The first set of surveys, initiated in 1966, consisted of four cohorts. These four groups are referred to as the 'older men', 'mature women', 'young men' and 'young women' cohorts of the NLS, and are known collectively as the 'original cohorts'. In 1979, a longitudinal study of a cohort of young men and women aged 14 to 22 was begun. This youth sample was called the National Longitudinal Survey of Youth 1979 (NLSY79). In 1986, the NLSY79 was expanded to include surveys of the children born to women in that cohort and called the NLSY79 Children. In 1997, the NLS programme was again expanded with a new cohort of young people aged 12 to 16 as of 31 December 1996. This new cohort is the NLSY97.
Web site: http://www.bls.gov/nls/

The BCS70 is a longitudinal cohort study which took as its subjects all those living in Great Britain who were born in the week between 5 and 11 April 1970. Its aim was to examine the patterns of maternity and obstetric care in Britain at the time. Since 1970 there have been four attempts to gather information from the full cohort: in 1975, 1980, 1986 and 1996 (a new survey of the whole cohort was planned for 1999). With each successive attempt, the scope of the enquiry has broadened from a strictly medical focus at birth, to encompass physical and educational development at the age of five, physical, educational and social development at the ages of 10 and 16, and physical, educational, social and economic development at the age of 26 (see Appendix 2 for details).
Web site: http://www.cls.ioe.ac.uk/Bcs70/bintro.htm

In Europe many prospective longitudinal studies were set up in the early 1980s (particularly in the period 1984–85). As already stated, it was not by chance that prospective studies started relatively late. From the 1970s on, all

advanced industrial societies had begun to undergo profound socio-economic changes which were accompanied by a process of polarisation – which increasingly distanced individuals who had access to the labour market from those who were relegated to the margins – and by a profound crisis in welfare systems which were proving unable to deal with an unstable system that was increasingly marked by long periods of unemployment (McFate, 1995: 3–4). The main purpose of HPSs is, in fact, to analyse income fluctuations and to describe and explain changes in the economic situation of the subjects studied, with aspects linked to monitoring poverty providing the background.

Apart from noting that Anglo-Saxon cultures clearly dominate in the field, it should also be remembered that longitudinal studies in Europe have in general been developing at two different speeds as the very high costs of such studies has made it difficult for less well-off countries to launch them. It is no coincidence that the countries of Northern Europe (Germany, Sweden, The Netherlands, Belgium, Great Britain) have been adopting a dynamic approach to the study of social phenomena for much longer than those in Southern Europe; indeed, there is still a severe lack of dynamic data available in the latter countries where there is no tradition of serious, in-depth, longitudinal research.

Notwithstanding this, the available longitudinal data, rather than cross-sectional data, can still make a fruitful contribution to understanding the way in which life conditions are evolving in these countries. As Yfantopoulos (1993) said, in countries like Greece, Spain, Portugal and Italy with their large agricultural and tourist sectors and 'invisible' market transactions, the informal component in both social and economic accounts is relatively large. Inevitably, the measurement error in cross-sections has influenced both empirical findings and, consequently, policy proposals. For example, in Greece, when researchers attempted to measure what quantity/value of their own goods a producer – in the agricultural sector – consumed, they met with problems of memory, of the validity of imputed prices and of seasonality. Thus, when well-developed panel methodologies are not used, it would appear that it becomes difficult to obtain reliable imputed values which can approximate the goods and services produced by informal economic activities, not only in the agricultural sector but also in the household sector.

Table 1.1 (pp. 20–3) shows the current availability of longitudinal data in Europe and in North America, according to their chronological order.

Notes

1 It is, however, true that longitudinal data are universally lacking for people with disabilities. This is an especially crucial omission for children, who change much more rapidly than

adults as regards disability. For example, the Survey of Income and Program Participation (SIPP) has collected data on children's disability since its inception, but none on children's Supplemental Security Income (SSI) participation, because SSI receipt is in the core set of questions designed solely for adults (this was remedied starting with the 1996 SIPP) (Eustis *et al.*, 1995).

2 As Trivellato wrote (1999), longitudinal data can, indeed, be obtained: (a) through repeated cross-sectional surveys which seek retrospective information about a fairly long period of time; (b) through panel surveys, which gather information from the same subjects who are interviewed, periodically, over a period of time; (c) by putting administrative records together and adding in any further information that can be drawn from census surveys (record linkages); (d) through a combination of all the three methods above.

3 Since then, there have been 12 censuses in Italy. Until 1931 the census was carried out regularly every 10 years (except for 1891, when it was not launched because of financial insolvency). However, on 6 November 1931 a law (no: 1503, art: 1) established that the census should be quinquennial. Five years later, in 1936, there was a census, but the next two (1941 and 1946) were missed because of the war and its aftermath. In 1951, census data began to be collected again and collections have continued, at regular 10-year intervals, ever since (Zajczyk, 1996).

4 At this point one could add the work by Tanner (1961, 1962, 1963) on adolescence and sexual development.

5 Life table analysis aims to study the passage of time before an event, or the time lapse between events. The basic idea is to sub-divide the observation period – starting from a specific point (e.g. beginning a job in a firm) – into a series of fixed short intervals of time (months or years). The probability within each interval, calculated for each subject studied over the chosen period, is used to calculate the probability of a terminal event (e.g. redundancy) which could take place within the chosen interval. These estimated probabilities are, in their turn, used to estimate the overall probability of the event which could take place at different points in time.

6 There are various types of biographical approach: (1) life history analysis; (2) study of the life-course; (3) study of life events (Olagnero and Saraceno, 1993). See Chapter 2 at p. 49 for further details.

7 At the end of the nineteenth century, in Europe, the Enlightenment and a romanticist faith in progress had given way to more disenchanted conceptions while, in North America, the pioneering ideal was still seen as a basis on which to found a new society. In this classic period of sociology many concepts were strongly influenced by the theory of evolution. North America, with its traditions based on those of a colony that had won independence, had no history or traditions of class struggle. This, along with the emergence of new problems (the violent repression of the emerging workers' movement, uncontrolled urban growth, millions of new immigrants often living in inhuman conditions), could in part explain why sociology was largely developed as a means of studying and resolving concrete problems: immigration, racial conflict, criminality, family breakdown, the isolation and 'ghettoisation' of new immigrants, and poverty.

8 For example:

- research by Thomas and Znaniecki (1918–20) on the breakdown of organisation in Polish families after mass-immigration from Europe to the United States. Using sources such as indirect documents (letters, diaries, etc.) and direct documents (specifically produced biographies) the two authors studied social change both within Polish society and among Poles who had emigrated to North America. The most important source was a collection of 754 letters either sent to, or received from, Polish immigrants in the US;

- the study of vagrancy carried out in the early 1920s by N. Anderson (1923);
- research carried out in the 1920s by a husband and wife team, Robert and Helen Lynd, on community life in an average North American town, called, unsurprisingly, Middletown (1929);
- the survey carried out in Boston by W.F. White on youth gangs (1943).

Table 1.1 Longitudinal studies in Europe and North America in chronological order

Name	*Country*	*Year*	*Time span*	*Unit of anlysis*	*Original sample*
National Child Development Study (NCDS)	Great Britain	1958	4–10 years (1965, 1969, 1974, 1981, 1991, 1999)	Cohorts	17,414 individuals born in the week 3–9 March 1958
National Longitudinal Surveys (NLS)	United States	1966	1–2 years	Cohorts	Around 5,000 individuals for each cohort: 'older men', 'mature women', 'young men' and 'young women'
Panel Study of Income Dynamics (PSID)	United States	1968	Yearly	Households	5,000
Swedish Level of Living Surveys (LNU)	Sweden	1968	7–8 years (1968, 1974, 1981, 1991)	Persons	6,000
1970 British Cohort Study (BCS70)	Great Britain	1970	5–10 years (1970, 1975, 1980, 1986, 1996)	Cohorts	17,198 individuals born in the week 5–11 April 1970
German Life History Study (GLHS)	Germany	1981–83	1–8 years	Cohorts	The West German Life History Study's data file contains information for 5,591 men and women of six different birth cohorts. In the East German Life History Study, 2,331 East Germans born in four different birth cohorts were interviewed

Survey of Income and Program Participation (SIPP)	United States	1983	4 months	Households	26,000
German Socio-Economic Panel (GSOEP)	Germany	1984	Yearly	Households	5,921
Household Market and Non-market Activities (HUS)	Sweden	1984	2–3 years (1984, 1986, 1988, 1991, 1993, 1996, 1998)	Households	2,619
Socio-Economic Panel Survey (SEP)	The Netherlands	1984	The 1984–89 period saw two waves per year carried out, in April and October of each year. In 1990 this was changed to one wave per year	Households	5,000
Panel Socio Économique 'Liewen zu Lëtzebuerg' (PSELL)	Luxembourg	1985	Yearly	Households	2,012
Socio-economic Survey – Panel des Ménages Lorrains (ESEML)	France	1985	Yearly ESEML ended in 1990, after the sixth wave had been collected	Households	2,092 The first wave 1985 was limited to a subsample of 715 households

(continued …)

Table 1.1 continued

Name	*Country*	*Year*	*Time span*	*Unit of anlysis*	*Original sample*
Belgian Socio-Economic Panel (SEP)	Belgium	1985	4–5 years (1985, 1988, 1992)	Households	6,471
Spanish Household Panel Survey – Encuesta Continua de Presupuestos Familiares (ECPF)	Spain	1985	3 months	Households	3,200
The OSA Labour Supply Panel	The Netherlands	1985	2 years	Households	4,020
The Polish Household Panel (PHP)	Poland	1987	Yearly	Households	2,100
The Irish Panel Study, now Living in Ireland Panel Surveys (LII)	Ireland	1987	Yearly after 1994 In 1994 (second wave), it became the Irish component of the ECHP survey	Households	3,321 (1987) 4,048 (1994)
The Swiss Household Panel 'Vivre en Suisse – Leben in der Schweiz' (SHP)	Switzerland	1999	Yearly	Households	5,074

The Bank of Italy Survey of Household Income and Wealth (SHIW)	Italy	Started in 1965: in 1989 a panel section was introduced	Yearly until 1987 (except for 1985) and every 2 years thereafter	Households	In 1989 about 15% of the sample (1,208 households) was obtained by re-interviewing families already interviewed in 1987
British Household Panel Study (BHPS)	Great Britain	1991	Yearly	Households	5,511
Panel Study of Belgian Households (PSBH)	Belgium	1992	Yearly	Households	4,439
The Hungarian Household Panel Study (HHP)	Hungary	1992	Yearly	Households	2,600
Survey of Labour and Income Dynamics (SLID)	Canada	1993	Yearly	Households	15,000
European Community Household Panel (ECHP)	European Community	1994	Yearly	Households	60,819
Longitudinal Study of Italian Families – Indagine Longitudinale sulle Famiglie Italiane (ILFI)	Italy	1997	2 years. The second wave (1999), has just finished; the third wave (2001) is, currently, being launched	Households	4,714

2 Longitudinal data

Characteristics and analytic advantages

There is no doubt about the heuristic potential of either prospective or retrospective longitudinal data. Indeed, such data make it possible to:

- analyse the duration of social phenomena;
- highlight differences or changes, between one period and another, in the values of one or more variables;
- identify *sleeper effects*, that is, connections between events and transitions that are widely separated in time because they took place in very different periods, as in the relation between childhood, adulthood and old age (Elder, 1985; Hakim, 1987). For example, the experience of old age has much to do with hardship in the adult years and one's responses to it: the same event or transition followed by different adaptations can lead to very different trajectories (Elder and Liker, 1982; Negri, 1990). Also caring at a young age has a significant effect on earnings and risk of poverty in later life as young carers are often absent from school and fail to gain even the basic qualifications (Olsen, 1996; Payne, 2001);
- describe subjects' intra-individual and inter-individual changes over time and monitor the magnitude and patterns of these changes;
- explain the changes in terms of certain other characteristics (these characteristics can be stable, such as gender, or unstable, that is, time-varying, such as income) (van der Kamp and Bijleveld, 1998: 3).

Longitudinal data also contribute to identifying the causes of social phenomena or at least they help to do this by allowing antecedents to be specified and consequences identified. The temporal ordering of events is often the closest we can get to causality: the structure of causality inherent in social processes may be reconstructed as a specific sequence of events leading to a certain state (Leisering and Walker, 1998b). More specifically, longitudinal studies not only allow the researcher to study the segment of the population which at different points in time finds itself caught within a specific situation, such as poverty or unemployment, but also, because of

their very nature, can be used in order to examine the flows, into and out of such a situation, thus opening up many paths for both causal analysis and for inference (Duncan and Kalton, 1987; Rose, 1993, 2000).

Indeed, longitudinal research gives a better insight than does cross-sectional research into the causal relations between variables. Three criteria are essential to establish whether or not there is a causal relation between two variables:

1 *Covariation*: the phenomena, or variables, of interest must be statistically associated (there must be a relation between X and Y). As each independent variable varies there must be an observable variation in the dependent variable too;
2 *Non spuriousness*: the relation must not be due to the effects of other variables, that is, must not disappear after controlling a third variable;
3 *Temporal order of events*: variations in cause (X) must intervene before variations in effect (Y): in a temporal sequence, the presumed cause must either precede or be simultaneous with the effect.

The first two criteria can, in principle, be tested using data from cross-sectional studies. Evidence relevant to the third criterion can usually only be obtained using longitudinal data that provide information about the temporal order of the designated 'cause' and 'effect' variables. As has been stated,

> In reality the panel also has important implications at the theoretical and conceptual level: it is one thing to follow individuals through time with the expectation that events will take place and quite another to reconstruct, using a selected group of persons, the relationship between what exists today and what has happened in the past.
>
> (Olagnero and Saraceno, 1993: 93)

There is also a fourth criterion, not usually mentioned (Taris, 2000: 3–4). Causal inference is theoretically driven. Causal inferences cannot be made directly from empirical designs: causal statements are based primarily on substantive hypotheses which the researcher develops. In other words, if we empirically observe that a variation in X is regularly followed by a variation in Y, while keeping all other possible causes of Y constant, then we have a strong empirical argument that corroborates the hypothesis that X is the cause of Y.

With longitudinal data it is also possible to develop causal theories that link individual dynamics with the dynamics of institutions and social structures (Gershuny, 1998, 2000), that is, which make it possible to fit the events studied both into individuals' biographies and into the family and social contexts they are part of, permitting in-depth analysis of social and

demographic processes in terms of both the choices and the determining factors that underlie different behaviours.[1]

Longitudinal data also allow us to construct more complicated behavioural models than purely cross-sectional or time-series data (Hsiao, 1986: 3; Davies and Dale, 1994: 4). More precisely, longitudinal data allow models to be constructed that are better able to take into account some of the complexities of the way in which people conduct their lives, i.e. models that allow improved control over the myriad of variables that are, inevitably, omitted from any analysis. Because of the complexity of human behaviour and because of our limited ability to model it, there is always considerable heterogeneity in the response variable, even among people with the same characteristics. For example, women with the same age, level of education and number of children will show considerable differences in their level of labour market participation. There are also other influences, which may differ between these women, which have not been measured and cannot be taken into account in the model. Omitting these variables may lead to misleading results, particularly if the variables omitted are correlated with one of the explanatory variables. Indeed, the effect of unobserved individual characteristics, which generally do not vary over time, can drastically undermine the results of analyses carried out on cross-sectional samples, because parameter estimates will be inconsistent. By using longitudinal information, one is better able to check for the effects of missing or unobserved variables, thus attenuating the effect of 'unobserved heterogeneity' – a key econometric problem that often arises in empirical studies – namely the assertion that the real reason one finds (or does not find) certain effects is because of omitted (mis-measured or not observed) variables that are correlated with explanatory variables. This problem can easily be overcome by exploiting the time invariance of the unobserved individual characteristics – a plausible assumption in many instances – and by the fact that repeated observations on the same individuals are available (Hsiao, 1985, 1986; Mátyás and Sevestre, 1996; Trivellato, 1999).[2]

Furthermore, the development of research projects which use longitudinal data serves to build a 'bridge' between quantitative and qualitative research traditions and encourages a reassessment of the concepts themselves of qualitative and quantitative research. The tendency to view the two research traditions as reflecting different epistemological positions and divergent paradigms has exaggerated the differences between them. Consequently, quantitative and qualitative research are often depicted as mutually exclusive models of the social process.

While qualitative research presents a process-oriented view of social life, lack of adequate data has forced many quantitative researchers to restrict themselves to carrying out static, cross-sectional studies with inference only

about process. Bryman (1988: 65–6) stated that there is an implicit longitudinal element built into much qualitative research: the general image that the qualitative researcher conveys about the social order is one of interconnection and change. Great emphasis is placed on social life as an interlocking series of events: this emphasis can be seen as a response to the qualitative researcher's concern to reflect the reality of everyday life which takes the form of a stream of interconnecting events. For example, the life history method is often depicted as being an important method of qualitative research because it entails the reconstruction of individual lives. Data sources may vary: from diaries to autobiographies, to unstructured interviews (life histories) in which the researcher/interviewer induces others to reflect at length about their lives and the changes and processes which underpin their experience (Bryman, 1988).

However, the social sciences are currently undergoing a period of rapid methodological development. Much of this progress has been stimulated by the growing recognition that analyses of social life based upon static, cross-sectional data are incomplete (Davies and Dale, 1994). Longitudinal surveys usually combine both extensive (quantitative) and intensive (qualitative) approaches. Life history surveys facilitate the construction of individual trajectories since they collect continuous information throughout the individual's life-course. Panel data trace individuals and households over time by gathering information about them at regular intervals. Moreover, they often include relevant retrospective information, so that the respondents have continuous records in key fields from the beginning of their lives. For these reasons, longitudinal data are well-suited to the statistical analysis of both social change and dynamic behaviour.

In the next paragraphs we will look more closely at the characteristics of each longitudinal design:

- repeated cross-sectional surveys
- panel design
  - consumer panels
  - prospective panels
  - rotating and split panels
  - cohort panels
  - linked or administrative panels
- event oriented design
- 'qualitative' longitudinal sources.

Repeated cross-sectional surveys

Surveys differ in the way in which they take time into account. The most common distinction is between *cross-sectional* and *longitudinal* studies.

As already described, a cross-sectional study analyses a cross-section of the population at a specific point in time. Details about an event/phenomenon are gathered once, and once only, for each subject or case studied. Consequently, cross-sectional studies offer an instant, but static, 'photograph' of the process being studied. Their one-off nature makes such studies easier to organise and cheap as well as giving them the advantage of immediacy, offering instant results. This is why they have always been the mainstay of both academic and market researchers.

However, as was argued in Chapter 1, cross-sectional studies are not the most suitable tools for the study of social change. Social scientists should be very careful when attempting to extrapolate longitudinal inferences on the basis of analyses of cross-sectional data as they have to, implicitly, assume that the process being studied is in some sort of equilibrium.[3]

Because of this, cross-sectional surveys are usually repeated twice or more, at different points in time, each time using a completely new sample. The samples include entirely different cases and any overlaps that may occur are so rare that they cannot be considered to be significant.[4] The term 'trend studies' is used for these repeated cross-sectional surveys (conducted at two or more occasions) on different samples. In order to ensure the comparability of the measurements across time, the same questionnaire should be used in all cross-sectional surveys. As Hagenaars (1991: 271) and Taris (2000) say, trend studies have some advantages over panel and cohort studies in that trend data are more readily available, can be analysed in a simpler way than cohort and panel data, and allow the detection of change at the *aggregate* level. The investigation of long-term social change in particular has to rely on trend rather than panel data. First of all, long-term panel data are scarce (see Table 2.1). Moreover, panel data suffer from attrition problems – that is, subsequent loss of membership due to non-contact, refusal to answer, failure to follow-up sample cases for other reasons, death, emigration (see Chapter 4 for details) – while cross-sectional surveys can be arranged into a long term trend design.

Cross-sectional data can be organised in two ways (Davies and Dale, 1994):

1 data gathered at the individual level (micro). Line-vectors (relative to cases) contain the same variables which are studied at different points in time. These can then be joined in order to create a single data *file* (a *pooled data file*). This increases the size of the sample and, also, makes it possible to insert a time dimension into the analysis;
2 data at the aggregate level (macro), where information about cases is compiled into tables in which time is considered to be the main independent variable. These aggregate data effectively bring together information, about the same population, but are gathered on a series of different occasions.

Example 2.1 Examples of repeated cross-sectional surveys

One example of a repeated cross-sectional survey (trend) is the Eurobarometer (EB), a unique programme of cross-national and cross-temporal comparative social research. Since the early 1970s representative national samples in all European Union (European Community) member countries have been simultaneously interviewed each spring and each autumn. The EB is designed to provide regular monitoring of social and political attitudes among EU publics. It has been conducted, every other year since 1973, in all countries which are part of the European Community (nine in 1973, 18 in 1999) on a sample of individuals, aged over 15, who are interviewed in their own homes. The regular sample size in standard EB surveys is 1,000 respondents per country, except in the United Kingdom (1,000 in Great Britain and 300 in Northern Ireland since EB 3) and Luxembourg (300 until EB 33, subsequently 500). They have included Greece since autumn 1980, Portugal and Spain since autumn 1985, and the former German Democratic Republic from autumn 1990 onwards (additional sample of 1,000 East Germans). In addition, an autonomous standard EB on selected sets of questions was established in Norway (1,000 individuals) in autumn 1991 and in Finland (1,000) in spring 1993. Austria (1,000) and Sweden (500) first joined in autumn 1994. The questionnaire is designed to periodically repeat the same questions and the survey aims to study attitudes, values and opinions in the political and social fields to enable comparisons to be made between the countries involved. Among the recurring themes are: attitudes to Europe, immigration, organisation of time on a daily basis, the condition of women, political opinions and materialist and post-materialist values (Corbetta, 1999).
Web site: http://www.gesis.org/en/data_service/eurobarometer/

A second important example is the General Household Survey (GHS). It is conducted by the Social Survey Division of the Office for National Statistics (ONS). This annual, multipurpose survey began in 1971 and data are available from 1973 onwards: it is based on a sample of around 10,000 private households in Great Britain. Interviews are conducted with everyone aged over 16 in the household (around 18,000 adults). The GHS offers researchers from a broad spectrum of disciplines opportunities to explore the relationships between income, housing, economic activity, family composition, fertility, education, leisure activities, drinking, smoking and health. The topics covered to date are listed each year in the GHS Annual Report, 'Living in Britain: Results from the GHS'. In addition to regular 'core' questions, certain subjects are covered periodically, such as family and household formation, health and related topics, use of social services by the elderly and participation in sports and leisure activities.
Web site: http://www.mimas.ac.uk/surveys/ghs/ghs_info.html

The Family Expenditure Survey (FES) is a comprehensive source of data on how families spend their money. It is a continuous household survey carried out by the ONS. Information is collected from a sample of around 7,000 households in the UK. It collects information about the income and expenditure of the household and, in addition, each individual spender over 16 is asked to complete an expenditure diary, listing every item bought over a period of two weeks and noting if a credit card was used to make the purchase (ONS, 1996). Thus, the FES provides detailed information about household expenditure on goods and services, with considerable detail in the categories used; information on income, including details about the sources of income; possession of consumer durables and cars; plus basic information on housing and many demographic and socio-economic variables which are mainly used for classification purposes.
Web sites: http://www.mimas.ac.uk/surveys/fes/
http://www.mimas.ac.uk/surveys/fes/fes_info.html
http://www.data-archive.ac.uk/findingData/fesAbstract.asp

Panel design

As already mentioned in Chapter 1, there are two types of true longitudinal survey: *prospective* and *retrospective*. The latter are based on historical accounts: subjects are asked to remember and to reconstruct aspects of their life-course; while the former gather information about events even as they are taking place. While trend studies analyse different subjects at different points in time, panel studies periodically gather information from the same subjects over the course of time (Arminger and Mueller, 1990; Engel and Reinecke, 1994).

The term 'panel data' covers a variety of data collection designs, but generally refers to the repeated observation of a set of fixed entities (people, firms, nation states) at fixed intervals (usually but not necessarily, annually) (Campbell, 1996). There are various basic types of panel.

Consumer panels

First, those which seek to ascertain the degree of stability or fluctuation of opinions and attitudes (usually surveys on political opinions or consumption). For example *consumer panels*, which are used in market research in order to keep track of changes in purchasing and consumption patterns in relation to a particular product (Sudman and Ferber, 1979). The participants in such panels provide the researcher with information on a regular basis about their level of consumption of particular brands of products (van de Pol, 1989). Data collection is at frequent intervals.

Household panel studies

The most representative prospective surveys, *household panel studies* or HPSs, are based on a probability sample of individuals/households, and seek to discover what happens/has happened to the same subjects over a certain period of time. The population from which the sample is drawn is made up of all the individuals resident/present in a given area or a subset of these individuals. HPSs are conducted using repeated interviews carried out at fixed intervals which could be anything from every two to three months to once a year (with some important exceptions: see Table 2.2 for details) the shorter the time interval, the easier it is for a relationship to develop with the household and the interviewees helping to ensure a high and constant percentage of response over time. Usually, the composition of the population is dynamic in two ways: it changes over time both in terms of entrants – through births and immigration, and leavers – through deaths and emigration. Second, its basic aggregate units – households – (which are also the sampling units of the HPSs) change continually, in the wake of events affecting family formation and dissolution (Trivellato, 1999). It is individuals, not households, who are followed over time: individuals are much more stable in a longitudinal context and so are easier to track and follow. Thus, if longitudinal surveys tell us about the dynamics of households, the data on these come from individuals who are related to their changing households and family contexts (Rose, 2000: 9). The clearest advantage these surveys offer is that they make it possible to study micro-social change. When individuals are studied over time it becomes possible to investigate the dynamics of both individual and family behaviours in the economic-social field and, also, the personal responses and adaptation strategies adopted in the face of previous circumstances and events. As already mentioned in Chapter 1, the most important examples of HPSs are, without doubt, the Panel Study of Income Dynamics (PSID) in the US, the Socio-economic Panel (GSOEP) in Germany and the British Household Panel Study (BHPS) in the UK. Proof of the growing importance attributed to HPSs can be found in the multiplicity of studies, set up in recent years, to examine and make comparisons, both *ex ante* and *ex post*, between longitudinal data. For example, in 1994, Eurostat launched a panel study – the European Community Household Panel (ECHP) – which extends over all the member countries of the European Union. Four projects have been set up which seek to increase the *ex post* comparability of prospective panel studies: the Panel Comparability Project (PACO), which aims to build up an archive of longitudinal data that can be compared at the supra-national level by drawing on various prospective longitudinal surveys currently under way in some European countries and in the US; the PSID-GSOEP Equivalent Data File, an attempt to compare GSOEP and PSID data; the European Panel Analysis

Group (EPAG) dataset and, lastly, the Consortium of Household Panels for European Socio-economic Research (CHER) project, whose aim is to develop a comparative database for longitudinal household studies by harmonising and integrating micro datasets from a large variety of panels (see Chapter 3 at pp. 63–9 for details).

Rotating panels and split panels

It is important to distinguish between *rotating panels* and *split panels* (Kish, 1986, 1987). The former are surveys in which a new group of individuals chosen via probability is added to the sample at each successive wave[5] to correct distortions which may have arisen within the sample between time t and time t_1 (e.g. one-sixth of the sample retire and are replaced by an equal number of employed persons). The idea is to keep samples of changing populations up-to-date. Sample size is controlled by stipulating the period of time any subject will be included in the survey, i.e. there is a limit on the time each subject will participate in the panel (e.g. two years). Such rotation serves both as a good method for maintaining the original characteristics of the sample and reduces the distortion which would otherwise be created by natural loss of subjects. This 'refreshing' of the sample has the advantage that subjects will develop 'survey boredom' less easily, that there will be fewer testing and learning effects, and that there will be less panel mortality. Thus, rotating panel surveys combine the features of both panel and repeated cross-section studies. Some important examples of studies which use rotation are: the Survey of Labour and Income Dynamics (SLID) in Canada; the Survey of Income and Program Participation (SIPP) in the United States; the Quarterly Labour Force Survey (QLFS) in the UK; the Household Budget Continuous Survey (Encuesta Continua de Presupuestos Familiares or ECPF) in Spain (Kalton and Lepkowski, 1985; Citro and Kalton, 1993).[6]

Example 2.2 The Survey of Labour and Income Dynamics (SLID), the Survey of Income and Program Participation (SIPP), the Labour Force Survey (LFS) and the Household Budget Continuous Survey (HBCS)

The SLID is one example of a rotating design. It is a longitudinal household survey conducted by Statistics Canada, designed to capture both the economic well-being of individuals and families over time and the determinants of their well-being. The first reference year of the survey was 1993. A second six-year panel of respondents was introduced in 1996 (wave 4), halfway through the life span of the first. A third panel started in 1999 (when the first panel ended) and a fourth will start in 2002 (wave 10). Each panel includes about 15,000 households (approximately 30,000 individuals aged 16 years and over). This pattern of rotating, overlapping panels will be continued with a new panel being selected

every three years. Individuals originally selected for the survey are interviewed once or twice per year for six years to collect information about their labour market experiences, income and family circumstances. In order to obtain complete information on families and to obtain cross-sectional data, people who live with the original respondents at any time during the six years are also interviewed.
Web sites: http://www.ssc.uwo.ca/sociology/longitudinal/Data.htm
Overview of the Survey of Labour and Income Dynamics (SLID) Philip Giles
http://www.statcan.ca/english/IPS/Data/75M0001XCB.htm

The SIPP is a continuous series of national panels, with sample size – a multistage-stratified sample of the US civilian non-institutionalised population – ranging from approximately 14,000 to 36,700 interviewed households. The duration of each panel ranges from two-and-a-half years to four years. The survey uses a four-month recall period, with approximately the same number of interviews being conducted in each month of the four-month period, for each wave. Interviewing for the first panel, the 1984 panel, began in October 1983 with a sample size of approximately 26,000 designated households. For the 1984–93 panels, a new panel of households was introduced each year in February. A new, four-year, 1996 panel was introduced in April 1996. The new 1996 panel consisted of 36,700 sample units (households). Households were to be interviewed 12 times from April 1996 through to March 2000. The survey collected data on source and amount of income, labour force information, programme participation and eligibility data, and general demographic characteristics, to measure the effectiveness of existing federal, state, and local programmes; to estimate future costs and coverage for government programmes, such as food stamps; and to provide improved statistics on the distribution of income in the country.
Web sites: http://www.sipp.census.gov/sipp/sipphome.htm
http://www.sipp.census.gov/sipp/sippov98.htm

The LFS is a repeated cross-sectional survey of households in the United Kingdom. It aims mainly to provide information on the UK labour market for international comparisons but also contain detailed questions of national interest. It is carried out by the Social Survey Division of the ONS in Great Britain and by the Central Survey Unit of the Department of Finance and Personnel in Northern Ireland, on behalf of the Department of Economic Development. The LFS was conducted biennially from 1973 to 1983 (for UK), and annually, with around 60,000 sampled households, from 1984 to 1991 for Great Britain, and from 1984 to 1994 for Northern Ireland. The QLFS has been conducted from the spring of 1992 for Britain and from the winter of 1994/95 for Northern Ireland. The sample size was increased to 60,000 households per quarter, which is equivalent to the size of the previous annual LFS. The QLFS, whose aim is to give stable, quarterly estimates of labour force, has a rotating quarterly panel design in which 80 per

cent of selected households are retained in the sample in successive quarters: every quarter is made up of five waves, each of which contains about 12,000 selected households. It was planned to interview households 12 times from April 1996 to March 2000. Accordingly, any quarter contains one wave receiving the first interview, one wave the second interview, and so on, and one the final/fifth interview.
Web sites: http://www.mimas.ac.uk/surveys/lfs/
http://www.mimas.ac.uk/surveys/qlfs/

The HBCS (or Encuesta Continua de Presupuestos Familiares – ECPF) was started by the Instituto Nacional de Estadistica (INE) in January 1985. It provides quarterly and annual information on the origin and amount of household incomes, and the way they are used for consumer spending on specific goods and services. The survey was targeted to 3,200 sample households. Half of the current sample (over 4,000 households) collaborates during one week per quarter by keeping a note in special notebooks of all the goods and services they have paid for during this period. However, because one week is too brief a time interval to be able to include the purchase of all the range of relevant goods and services of consumption, information is also asked for, through interviewing the totality of the sample (over 8,000 households), about purchases regularly carried out at intervals greater than a week. Every quarter, one-eighth of the sample is renewed, so every household collaborates for a maximum of eight quarters.
Web sites: http://www.ine.es/welcoing.htm
http://www.ine.es/dacoin/dacoinme/inotecpf.htm

Split panels are 'classic' panels which include a rotating sample that is interviewed alongside another sample of the long-term panel members who are being followed over time. The rotating sample is interviewed once only and never again and serves as a control group as they are not exposed to the potential effects of participating in the survey (attrition and conditioning). A panel study is, therefore, combined with a repeated cross-sectional study (van de Pol, 1989), by flanking one-off independent samples with the long-term sample. The British Social Attitudes Survey (BSA) is an example of a split panel survey.

Cohort panels

Cohort analyses are similar to panel studies except the same individuals are interviewed in each period. As already described, in cohort studies only a random sample of the individuals who experienced the same life-event within the same time interval is followed over time. Usually a researcher will choose one or more birth cohorts and administer a questionnaire to a sample drawn from within that group: thus longitudinal analysis is used on groups of the

same age and a number of generations are followed, over time, throughout their life-courses.

A cohort study may not necessarily start with the birth of the interviewees: a good example of this are the National Longitudinal Surveys (NLS) which, retrospectively, gathered the work histories of two cohorts of men and women aged between 30–44 years in 1967 and those between 24–37 years in 1978 (Centre for Human Resource Research, 1981). Another example is the series of cohort studies funded by the Medical Research Council and based in the Medical Sociology Unit at Glasgow. The study sampled cohorts born in 1931, 1951 and 1971 in the area of Glasgow (Davies and Dale, 1994). Unlike HPSs, where there is a dynamic population – which changes over time because of birth, deaths, immigration, etc. and where family organisation may change because of divorce, re-marriage, a new marriage or children leaving home – the main characteristic of this type of research is that a cohort is closed against new entries because such entries are, by definition, impossible (Ghellini and Trivellato, 1996). An example of this is panel studies on scholastic career (and/or on the transition from school to an active working life), where the event-origin used to identify the cohort is that of being present in (or entering or leaving) a given class in a given school year. This type of study is used to investigate the particular experiences of specific groups of people – which it does by analysing changes over the long term: indeed subjects are usually re-interviewed only every five years.

If, in every specific generation, the same people are followed over time then the cohort study will be composed of a series of panel studies; however if, for each observation, a sample is chosen from within each generation the cohort study will consist of a series of trend studies. Cohort studies can be either prospective or retrospective. The former usually study one or more cohorts, at successive intervals, over a period of time, while the latter gather retrospective information about just one cohort at a time and may thus be made up of more than one study. Because of this, retrospective studies may evince, simultaneously, both cross-sectional features (samples are only interviewed once) and prospective panel features (they offer information about the life histories of the interviewees). Examples of the first group are the National Child Development Study (NCDS) and the 1970 British Cohort Study (BCS70), both British, and the series of NLS carried out in the United States (see Appendix 2 for further details). The German Life History Study (GLHS) is a good example of the second group.[7]

The underlying idea behind any cohort study is that long-term social change must be interpreted within the context of generational change. By following one generation throughout its entire life-course, the consequences of growth, maturity and ageing are rendered visible. Furthermore, it also becomes possible to investigate the influence of a variety of events that take

place over the course of time and, likewise, to understand whether a specific event has influenced an entire generation in the same way (Hagenaars, 1990). Consequently, cohort studies are particularly suitable when studying populations that are subject to radical changes (Olagnero and Saraceno, 1993).

To sum up, cohort studies could be considered to be a specific form of panel study where both the process of rotation and of the substitution of one generation for another is explicitly taken into account and where the cohort effect, within a specific population, is duly corrected. There are indeed three types of changes in attitudes or behaviour of cohorts (see, among others, Glenn, 1977; DeGraaf, 1999). The first type of change may be a product of the age of the individual concerned, that is, is associated with changes in age (age effect). Changes of the second type – called cohort or generation effects – are associated with the time when the individual was born, and concern all events that one generation experienced and other generations did not. Finally, period effects concern those events which affect all generations equally and simultaneously, that is, the period at which the data were collected.

Example 2.3 Age, period and cohort effects

Inserting the variable 'time' into analyses has at least three different effects associated with three temporal dimensions and three levels of experience and change.

- *Age effects* concern all events associated with changes in age. Here chronological age is taken as one indicator of levels of maturity and of both physical and psychological skills. The specific effects of age will, obviously, vary from one age to another but are the same for all those who are part of a specific age group (Saraceno, 1986).
- *Period effects* concern those events which affect all cohorts equally and simultaneously. In the limited sense, the period effect refers to the time at which the observation is carried out. In practice, the concept is used as an indicator of the effects of events which affect all generations equally and simultaneously and which will have taken place during the observation period or between two consecutive observations, e.g. the long-term influences of processes such as industrialisation or urbanisation, etc.). Individuals who are born in different historical epochs will come into contact with social circumstances which will affect, modify, the passages connected to age and the phases of life. These effects vary over the course of time but will, however, be the same for all subjects at any one particular point in time.
- *Cohort effects* are associated with year of birth and concern all those events that one cohort has experienced and others have not. They are often interpreted as a special interaction between age and period effects: they interpret growth within specific historical conditions. A cohort can be defined as a

group, or set of persons, who have experienced the same event-origin within a given interval of time: birth, first marriage, reaching the age-of-consent, etc. If the distinguishing event is birth, one speaks of a birth cohort (or of a marriage, work, graduation, etc. cohort, as the case may be) and this cohort is formed by all those who were born during the same period of time (e.g. everyone born in March 1970). A more restricted conception of cohort – which is better suited to a life-course approach – is that a cohort is a group of individuals who began their life-course during the same interval of time (Billari, 1998). Cohort effects are the same for all the individuals born within a specific, predefined period of time but will vary from interviewee to interviewee. Cohorts not only differ one from another, but are also not homogeneous internally. Given that the members of one cohort will be of different genders, health conditions, social classes, etc., they may not only have different life-course models but, also, they may be influenced differently by the same historical events. Elder (1974) demonstrated this in a study of people who entered adolescence during the Great Depression in the 1930s when he showed that the same events may even have the opposite effects on the male and female cohorts involved (Elder, 1974; Saraceno, 1986; Hagenaars, 1990).[8]

The term generation is often used in order to clarify the inter-connection between age, period and cohort effects. Indeed, generation expresses the socio-cultural changes which highlight the historical aspects of cohorts. Membership of a generation is usually defined as follows: being born within the same time period, undergoing certain, more or less similar, social, cultural and psychological experiences; being exposed to analogous primary and secondary socialisation processes (Gallino, 1993). In Mannheim's view (1952), a generation is not merely a birth cohort: historical events (especially if they occurred during the 'formative' period, i.e. around 15 years of age) may determine a whole generation's capacity for cultural elaboration, stimulate a common world view and, consequently, encourage the development of the consciousness of being a socio-cultural entity.

These three effects are often interlinked, so unless we can assume constancy for one or two of them we can never be certain which one we are observing in the longitudinal data. The problem is that these effects cannot be identified separately, since they are linearly dependent: cohort equals period minus age. This equality is known as the identification problem (van der Kamp and Bijleveld, 1998; De Graaf, 1999): knowing any two, fixes the value of the third. For instance, knowing period and age fixes the value of cohort.

So researchers need to assess the significance of these effects and to exercise a degree of control over them through their research design. To assess the extent of the cohort effect and to check up on it, we need to collect data from individuals of the same age but born at different points in time, that is, in different cohorts. To assess and check on the age effect, we need to collect data from individuals

of different ages in the same period. To assess and check on the period effect, we need to collect data from individuals of the same age at different periods (Bynner, 1996). From the point of view of the techniques, there is no solution to the identification problem, only strategies for trying to deal with it. Sometimes it is possible to fix the value of the regression coefficient – a regression line is a good way to describe and summarise the linear relationship between two or more variables – for one of the effects, usually to zero. For instance, in a study of political preferences, one might assume that the age effect was zero and all changes over time were due to period and cohort effects. One could then use 'period' and 'cohort' as independents in a regression in which political preference was a dependent, but results would be invalid if the assumption that there was no age effect was an untrue assumption. One way of dealing with such assumptions is to run three regressions, each time fixing one of the effects (age, cohort, period) to zero, then examining the resulting coefficients to assess whether, on the basis of external information, all three models seemed plausible. One may find, for instance, that a regression coefficient approaches zero for one of the models, yet one has reason to believe that the effect for that coefficient does indeed exist, meaning that that model is not plausible (Firebaugh, 1997).

Linked or administrative panels

Linked or administrative panels are derived, as by-product, from data collected as part of public administration processes (census or administrative data). The value added may come from joining disparate data sources, e.g. registration data attached to the census information. In these cases, data items which are not collected primarily for panel purposes are linked together using unique personal identifiers (the combination of name, birthdate and place of birth is normally enough to identify individuals and enable linkage of administrative and/or other records). One good example of such panels is the ONS Longitudinal Study (LS), organised by the ONS: it is based on the census and vital events data (births, cancer, deaths) collected for a 1 per cent sample of the population of England and Wales (approximately 500,000 individuals at any one point in time). The LS study was established in the early 1970s. While the original LS sample took all people who gave one of four dates of birth at the 1971 census, the study has been continuously updated to include new births and immigrants born on one of these dates: this distinguishes the LS from other longitudinal studies where the sample is selected at one point in time (CLS, 1999). Another example is the Echantillon Démographique Permanent (EDP), launched in 1968, in France: in this case too, the study has involved more than 1 per cent of the census population. A third example is the Turin Longitudinal Study (TLS) in Italy: it is a longitudinal study containing linked census data for all persons who were resident in Turin at one or more

of the last three decennial censuses (1971, 1981 and 1991). This archive uses record linkage to bring together information relating to events (such as mortality, migration or ill health) which affect, or have affected, Turin residents. The TLS has been widely used for investigating health inequalities, including infant and adolescent mortality, and drug-related causes of death (Office for National Statistics and Centre for Longitudinal Studies, 2001). Administrative panels are particularly widespread in the Scandinavian countries: the Finnish Longitudinal Study (launched in 1971), where the whole resident population is being studied (Bynner, 1996); the Integrated Database for Labour Market Research (IDA) in Denmark; the Longitudinal Individual Data for Sweden (LINDA) and the Swedish Income Panel (SWIP) should be mentioned.

Example 2.4 Examples of linked panels

The IDA was primarily set up to make data available to labour market researchers. The database contains information on labour market conditions for persons and establishments (Danmarks Statistik, 1991). In this way, the IDA can be used for analyses on the basis of both the demand side (establishments) and the supply side (persons). The database is longitudinal: it contains annual information about the entire Danish population and all companies with employees for the period 1980–98. There are more than 200 variables in the database, including a vast number of background variables related to the population. Data are drawn from a wide range of registers,[9] where the most important are information on recruitment during the year from tax registers and information on unemployment, which is also obtained from an administrative register (Danmarks Statistik, 1994; Leth-Sørensen, 1997).
Web site: http://www2.dst.dk/internet/varedeklaration/en/V01013.htm

LINDA is a register-based longitudinal dataset. It consists of a large panel of individuals, and their household members, which is representative of the Swedish population. LINDA collects information on 300,000 individuals annually. Attached to LINDA there is a specific, non-overlapping sample of immigrants: this particular sample has the same design and covers the same period as the overall sample. The core registers consist of the income registers – available annually for the period 1968–97 – and population census data – available every fifth year from 1960 to 1990. All variables in these registers are included in the database. The database is updated annually: for each year, information on all household members of the sampled individuals is added to the dataset. Household members are included in the sample as long as they belong to a sampled household. LINDA was developed through a jointly funded effort by the Department of Economics at Uppsala University, the National Social Insurance Board (RFV), Statistics Sweden and the Ministries of Finance and Labour.
Web site: http://www.ehl.lu.se/database/linda.htm

The SWIP was originally set up at the beginning of the 1990s to study how immigrants are assimilated into the Swedish labour market (Gustafsson, 1997). It is made up of large samples of both foreign-born and Swedish-born persons. Income information from registers has been gathered for a period of 25 years. Samples are taken from a register of the total population (RTB) kept by Statistics Sweden (asylum seekers waiting for a residence permit are excluded). At present, the SWIP has information from the registers for each identity (person sampled, present spouse, mother and father as well as present spouses of mother and father) for a period of 25 years, 1968–92. Information from the income-registers covers demographic variables, education and different variables measuring income. A 1 per cent sample of native-born persons (about 77,000 individuals) was taken from the register for 1978, as well as a 10 per cent sample of foreign-born persons (about 60,000 individuals). A further 10 per cent of the people immigrating each year from 1979 until 1992 was also taken (sample sizes vary between 3,000–7,000 individuals). An update is planned: it will consist of taking new samples for people immigrating to Sweden in 1994, 1995 and 1996 and adding supplements for persons born in Sweden after 1978 (Gustafsson, 1997).

Without doubt, administrative panels offer the least intrusive method of collecting longitudinal data. Moreover, the datasets obtained are large, thus sampling errors are small and they are also cheap. Another of the advantages of using register information as primary data in connection with surveys covering a longer period of time, is that the effects of oblivion or memory can be reduced (Leth-Sørensen, 1997). However, they do have some clear disadvantages. Above all, they can only offer a very small variety of information, data which have often been collected with long intervals of time elapsing between one collection and the next (as in the case of census data), furthermore such data often pose comparability problems. One common problem with register data is comparability between the years covered. For example, in SWIP a fundamental problem for details of earnings and other variables obtained from tax records is changes in the tax code. During the period covered by the panel there were two important changes in tax codes. In 1974, a number of transfers from the public sector (compensations for sickness, unemployment compensation, etc.) became subject to income tax. The tax reform at the beginning of the 1990s broadened the tax base, and therefore income recorded in 1991 and after is not strictly comparable with income recorded earlier (Gustafsson, 1997).

Furthermore, the analytical possibilities such panels offer are limited to those issues which correspond to the bureaucratic concerns of the administrators who collect the data (Gershuny and Buck, 2000). Lastly, these panel studies are frequently impeded by laws concerning data protection, which

may make it difficult to obtain access to such data (Buck *et al.*, 1994; Bynner, 1996).

All these designs can, of course, be fruitfully combined. As already discussed in Chapter 1, the BHPS collected complete life and work history records for all its respondents using retrospection: indeed one of the options now beginning to be discussed for the 'missing' mid-1980s British birth cohort is a combination of administrative and other records with a new interview sample of adolescents (Gershuny and Buck, 2000).

The analytical advantages panel studies offer surveys based on samples – when compared with those offered by repeated cross-sectional studies or by a single retrospective study – have often been highlighted in the literature.[10] Panel studies are indispensable when one wishes to:

- adequately describe and analyse processes of mobility/inertia, that is, make a distinction between the transitory characteristics and the enduring characteristics of a phenomenon (e.g. poverty);
- describe flows (that is, transitions between states), which is essential for any analysis of mobility from one state to another (e.g. in the labour market or in social classes);
- conduct studies on the inter-generational consequences of phenomena such as poverty or dependence on public assistance and welfare programmes, consequences which would be hard to reveal through retrospective cross-sectional studies because of the unreliability of people's memories (Duncan, 1992; Ghellini, 1994; Ghellini and Trivellato, 1996).[11]

Indeed, the advantages of analyses carried out using prospective data are not hard to see. While cross-sectional studies do not reveal whether any changes that show up should be attributed to new individuals entering or to a real change in behaviour, panel studies resolve this problem as they offer researchers the opportunity of re-interviewing the same subjects again and again. This makes them an indispensable tool for the analysis of social change, of evolution in behaviours and of both individual and family change: longitudinal prospective studies allow the life-course of one individual to be followed over time. Thus, statistics are used here to examine the relations between the distribution of one variable (identified at the household or the individual level) which refers to one time and the distribution of the same variable or of other variables at a different time.[12]

The prospective approach also makes it possible to discern the dynamics of behaviours that may be discontinuous or difficult to analyse: the results of analyses carried out using panels have shown that changes in the lives of families and/or individuals are considerably greater than would appear from just taking into account single snapshots (Stouffer, 1950; Kasprzyk *et al.*, 1989; Dale and Davies, 1994a).

Panel studies also offer the opportunity of analysing change in a way which takes into consideration different *dimensions of time*. For example, the GSOEP contains questions which measure time in several ways (Frick, 1998).

- Single retrospective questions on certain events in the past (past time): e.g. how often have you changed your job during the last 10 years?
- Retrospective life-event history since the age of 15 (past time): e.g. employment or marital history.
- Monthly calendar on income and labour market related issues (past time): e.g. employment status January to December last year.
- Questions concerning a period of time (past time): e.g. demographic changes since the last interview (such as marriage or death of spouse).
- Questions about a point of time (present time): e.g. current employment status or current levels of satisfaction.
- Questions concerning future prospects (future): e.g. satisfaction with life five years from now, or job expectations.

Finally, another major advantage panel studies have over cross-sectional research designs is that they offer the possibility of performing an analysis of causal interrelationships among variables: panel data offer multiple ways of strengthening the causal inference process (Stouffer, 1950; Bulmer, 1983; Finkel, 1995). As Engel and Reinecke (1996: 8) wrote, one great virtue of panel data analysis is its ability to subject causal propositions to rigorous empirical examinations. Because for each unit of analysis, panel data place not only one but at least two or more repeated observations at the researcher's disposal – and this in definite time order – it appears much more reasonable, than in cross-sectional research, to infer ongoing processes. Since these observations are not collected retrospectively, as is often the case in event history analysis, memory and possible re-evaluation of past experience cannot distort the data.

Event oriented design (event history data)

However, repeated cross-sectional and longitudinal prospective data do have one important element in common which constitutes an important limitation for both: they are gathered at discrete points in time (e.g. every six months or annually). Indeed, any analysis of the evolution of many types of social phenomena really requires continuous (in time) investigation of discrete events in order to permit study both of the sequence of the events that have taken place and of the precise intervals which may have elapsed between one event and another – such information is crucial if one is to understand the development of a life-course and the way in which events and processes are interrelated.

Because events are defined in terms of changes over time, it is usually accepted that the best way to study them together with their causes is to

gather duration data or event history data, that is, to identify vectors which record what has happened to a sample of individuals, or a collective, together with precise information about the point in time when these events took place.

As already described, duration data are typically collected retrospectively through life history studies – which generally cover the whole life-course of individuals – or through the use of event histories, gathered using either prospective panels or cohort studies. In the former case, a sample of respondents are interviewed about aspects of their lives: e.g. they may be asked about all jobs and spells of unemployment they have experienced since leaving school. In the latter case, members of a sample are tracked over time and questioned every so often about what has happened to them: e.g. about all the important events that have affected household members since the last interview (Gilbert, 1993: 168).

Detailed information about each episode is collected: the duration of the event, the origin state and the destination state – one example could be the event 'first marriage': every individual who marries for the first time (*origin state* or *initial event*) starts off an episode which will only finish with the transition into the state of 'no-longer married' (*destination state* or *terminal event*) (Blossfeld and Rohwer, 1995).[13]

Example 2.5 Examples of retrospective questions

The measurement of event histories is generally based on the following types of retrospective questions:

1 Has the initial event ever occurred?
2 When did it occur?
3 Has the terminal event occurred?
4 When did it occur?

Time in (2) and (4) can be measured in several ways, e.g. age at occurrence, date of occurrence or time between occurrence and survey. Any of the measurements may be recorded continuously or grouped into intervals (Skinner, 2000: 121).

Furthermore, such studies often collect information relating to repeated episodes/events (consecutive jobs, unions, separations, births …) which take place during and alongside parallel processes (work, matrimonial, family histories, etc.) and at different levels (micro, meso and macro: e.g. individual work history, history of the firm in which the individual is employed, structural changes in the labour market). The underlying idea, or principle, is that an individual's life-course can only be understood if or when it is placed into the context of the trajectories of his/her social life. Because the changes

which exist or take place at a 'macro' level will potentially affect the life-course of an individual, then this life-course should not be isolated from the 'situation' in which it is set (Mayer, 1990). In Abrams' view (1982: 360):

> certainly, the lives of individuals are unique but their uniqueness does not depend on personal, intangible factors rather it is based on the diversity of moves that individuals, historically placed within historically determined social worlds, can make.

Wright Mills, too, stated (1959: 167):

> the biographies of men and women, of the different individuals that they become, cannot be understood if they are not considered in relation to those social structures within which and within whose context their daily lives are organised.[14]

Last, in Mayer's opinion (1990):

> Life-courses are shaped by a large number of inputs: specific structures offering political and economic opportunities; ideas shaped by the culture; norms that stipulate legal age for certain activities; sequence of positions and institutional passages; socialisation processes and selection mechanisms.[15]

In other words, these data make it possible to analyse developments within the institutional, cultural and social context in which an individual's life-course is unfolding because, by focusing on events and transitions in individual lives, the interaction between action and structures can be closely observed.

Thus, in an event oriented matrix each line vector corresponds to the duration of one state or episode: e.g. it could express a work/job episode (first job, second job, third job). If only one episode is considered for each case (e.g. the birth of the first child or the first marriage), then the number of vectors will correspond to the number of cases examined. If, however, these are repeated and/or parallel episodes, the number of which may vary greatly from one individual to another, the sum of the episodes that characterise *each* individual life-course represents the total of line vectors in the data matrix.[16]

One good example of a study oriented towards events is the already cited German Life History Study (GLHS), which is made up of a set of retrospective cohort studies that seek to gather detailed information both about events in the lives of the subjects involved and about their most important activities (see Chapter 3 at pp. 59–62 for details). The study is made up of diverse studies (12 in all) of cohort samples drawn from the population of Germany. These cohorts were not followed over time, but were contacted *just once* during the data-gathering activities. The groups were chosen in such a way that the transition phase between school and work coincided with

periods that were particularly important from the historical point of view: the immediate post-Second World War period; in a period of fast economic growth (boom); in a period of expansion within the welfare state and during a period of contraction in the economy (slump). The fundamental hypothesis underlying this study was that specific historical conditions would have had an equally specific impact on the working lives of those interviewed. As well as information about education and work, the GLHS also offers information about other important aspects of individual life: cultural background, family and residential history, etc. (Blossfeld *et al.* 1989: 17–25).[17]

One further example is the UK 1980 Women and Employment Survey (WES), which collected very detailed work histories from a nationwide sample of more than 5,000 working women aged between 16 and 59 years living in Great Britain (Martin and Roberts, 1984a).

Example 2.6 WES

The WES was commissioned by the Department of Employment and carried out jointly by the Office for Population Censuses and Surveys and the Department of Employment. The fieldwork took place in 1980 and was carried out by Social Survey Division of OPCS. The survey covered a nationally representative probability sample of 5,588 women in Great Britain aged 16–59 and the husbands of 799 of the married women. The response rate to the main survey was 83 per cent. Interviewers carried out short screening interviews at a sample of 9,944 addresses in order to identify women within the eligible age range who were then approached for the full interview.

The main aims of the survey were to establish what factors determine whether or not women are in paid work and to identify the degree to which domestic factors and the sexual division of labour shape women's lifetime labour market involvement; and to collect full information about the work they do, their pay and conditions of employment, as well as how they behave in the labour market when they leave jobs or look for work. The study also set out to determine the importance of work to women and their job priorities.

An important and innovative feature of the survey was the collection of detailed work histories covering the whole of women's working lives since leaving full-time education and detailed histories of other vital events such as the births of children, which were likely to have consequences for women's labour market behaviour. Major topics covered by WES are: current economic activity; details of current job; child care arrangements; attitudes to work; education and training; future employment plans; reasons for not working; job search activities; work and life histories; details of husbands' work and attitudes; general attitudes to employment and gender roles; financial circumstances (Martin and Roberts, 1984a, 1984b).

Web site: http://qb.soc.surrey.ac.uk/surveys/wes/wesintro.htm

---

One should also mention the Indagine Longitudinale sulle Famiglie Italiane (ILFI) (Longitudinal Survey of Italian Families): a prospective panel study with a retrospective first wave. This survey has set itself two main goals: to collect information about the situation of a sample of Italian families (household composition, income sources and levels, demographic and social characteristics of each nucleus) and to study social change by gathering both retrospective and prospective information about each adult – 18 years and over – who is a member of a household included in the sample. The survey seeks to reconstruct the life history of each household member (from birth up to the last wave of interviews – planned for 2005) in relation to their geographical and residential mobility, level of education and training, work history, social origins and, also, to changes in the composition of the household itself. In 1997, during the first wave of interviews, retrospective information was gathered about all the important events that had affected, or happened to, members of the sample from birth to the date of the interview. In each subsequent interview (the second prospective wave was collected in 1999), this information is updated in order to record all the important events that have affected household members since the last interview (Schizzerotto, 1999).

Two good examples of prospective studies that retrospectively investigate the life of the interviewee are the BHPS and the GSOEP. The BHPS has taken the opportunity (over the first three waves) to get a very good picture of respondents' lives by asking for life-time retrospective work-histories, and marital and fertility histories, hence investigating and illuminating vital areas of the lives of those who make up a representative sample of the households of Britain. In other words, quantitative and qualitative pieces of information are being linked together (Table 2.1).

The GSOEP includes two calendars in the core questionnaires:

1 an activity calendar that, on a monthly basis, records participation in schooling, vocational education, military service, full-time and part-time employment, unemployment, homemaking and retirement for the previous year (Table 2.2);
2 an income calendar where respondents indicate, also on a monthly basis, whether they have received income from various sources in the past year and the average monthly amount received from each source (Burkhauser, 1991).

Moreover, the GSOEP provides spell-oriented data on 12 different kinds of labour-market involvement, defining the beginning, end and censoring status of any period of work, i.e. full-time work, part-time work, or unemployment. Additionally, the database contains such data on the periods in which a person received different types of income (such as income from employ-

ment, pensions, unemployment benefits). Although over the course of time the absolute number of observations (households and individuals) has steadily decreased from a cross-sectional perspective, the number of events and/or periods covered by the data gathered has been increasing wave by wave: e.g. the return of foreigners to their home countries (re-migration) or births and deaths (fertility and mortality) (Merz and Rauberger, 1993; Frick, 1998).

What advantages do event oriented data offer that other types of longitudinal data do not provide?

- Above all, this type of data is the only type that makes it possible to investigate not only changes in state but also to discover exactly when, in time, these changes took/take place: i.e. these data make detailed reconstruction of all the phases of familial, educational and work histories possible and, also, set these histories into a precise historical context.
- Prospective panels do not gather information about what happens between one wave and another; however, events and changes in the variables studied can be continuously monitored and recorded using duration data.
- To sum up, these studies make it easier to construct individual trajectories because they gather information throughout a life-course. This makes it possible to study life events – e.g. passages from one state to another – within an individual's life on the basis of time vectors which can be dealt with using statistical procedures. Analysis of the history of life events not only highlights but also quantifies the extent of the interweaving of different times, underlining their length, putting the sequence of events into the right order, noting recurrences and measuring intervals (Olagnero and Saraceno, 1993; Mayer and Tuma, 1994).

'Qualitative' longitudinal sources

Longitudinal research does not only use data gathered through surveys on samples or that derived from mixing data from samples with information from censuses or administrative records. Another important longitudinal source can be found in the collection techniques used in biographical analyses. This type of analysis, which takes many forms (life history, study of life-courses and life events), is increasingly showing both theoretical and epistemological autonomy and developing a considerable repertory of themes. These spaces are more or less the same as those where it would in any case be good to use non-standard techniques, i.e. where the interview has to dig deeper than would be possible if structured, formal methods were to be used: the paths of physical and psychological distress; situations of economic, social and cultural marginalisation; deviance; mobility and career

Table 2.1 Life history calendar in the BHPS

Marital/fertility dates																							
1910–1929 Year 19..		10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Age left school	x	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Self-employed	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Full-time employee	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Part-time employee	3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Unemployed	4	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Retired	5	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Looking after family	6	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Other	7	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
1930–1949 Year 19..		30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49
Age left school	x	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Self-employed	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Full-time employee	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Part-time employee	3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Unemployed	4	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Retired	5	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Looking after family	6	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Other	7	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
1950–1969 Year 19..		50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
Age left school	x	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Self-employed	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Full-time employee	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Part-time employee	3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Unemployed	4	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Retired	5	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Looking after family	6	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Other	7	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
1970–1989 Year 19..		70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89
Age left school	x	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Self-employed	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Full-time employed	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Part-time employee	3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Unemployed	4	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Retired	5	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Looking after family	6	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Other	7	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
1992+ Year 19..		90	91	92																	
Age left school	x	–	–	–																	
Self-employed	1	–	–	–																	
Full-time employee	2	–	–	–																	
Part-time employee	3	–	–	–																	
Uneemployed	4	–	–	–																	
Retired	5	–	–	–																	
Looking after family	6	–	–	–																	
Other	7	–	–	–																	

Source: BHPS documentation, distributed to users on CD-ROM

Table 2.2 Activity calendar in the GSOEP

Activities	*Months*											
	J	*F*	*M*	*A*	*M*	*J*	*J*	*A*	*S*	*O*	*N*	*D*
Full-time employment (or job-creation measure)	–	–	–	–	–	–	–	–	–	–	–	–
Short-time work or waiting list	–	–	–	–	–	–	–	–	–	–	–	–
Part-time work or occasionally employed	–	–	–	–	–	–	–	–	–	–	–	–
Vocational training, education, retraining	–	–	–	–	–	–	–	–	–	–	–	–
Registered unemployed	–	–	–	–	–	–	–	–	–	–	–	–
Retired, early retirement	–	–	–	–	–	–	–	–	–	–	–	–
Maternity leave	–	–	–	–	–	–	–	–	–	–	–	–
In school, college	–	–	–	–	–	–	–	–	–	–	–	–
Military/civilian service	–	–	–	–	–	–	–	–	–	–	–	–
Housewife/househusband	–	–	–	–	–	–	–	–	–	–	–	–
Other	–	–	–	–	–	–	–	–	–	–	–	–

Source: GSOEP documentation, distributed to users on CD-ROM

events; changes in role especially in relation to gender characteristics; and transitions and changes in status, particularly in relation to age. Indeed, the biographical method aims to unravel the subjective dimension of time, the perceptions, orientations and self-interpretation that people develop during the course of their lives. Thus, this type of research demands a high degree of creativity from its practitioners (Walker and Leisering, 1998: 2829).

Biographical material may be collected either directly, i.e. through structured and/or in-depth interviews, or indirectly (Olagnero and Saraceno, 1993: 90–102; Corbetta, 1999: 438–43).

Different types of interview can be used for the direct collection of information:

- Relatively structured biographical interviews which aim to reconstruct events and behaviours: e.g. marriage and birth, relations with institutions (school or work), moving around the territory, consumption and saving behaviour. Such interviews may reconstruct the events experienced either entirely through retrospective investigation or through repeated interviews over a period of time. The problem with this type of interview, however, lies in the fact that the researcher has to trust entirely in the reliability of the subject's memory to discover how, and under what conditions, a previous situation started and then developed. The degree of error will also depend on how well or badly the interviews, and questionnaires, are structured. There are techniques that can be used to minimise distortion: e.g. it is better to start an interview with something that is based entirely on recall and then go on to more complex questions which require opinions. A retrospective interview is fairly reliable when

it is dealing with crucial past events or transitions, such as those concerning work, marriage, changes in family composition, maternity, etc. Results are less trustworthy when one wants to find out about the precise and detailed results of short-term life plans, or about more complex matters, such as economic strategies and behaviour.

- Semi-structured or unstructured interviews, suitable for discerning the cultural/symbolic level of the discourse, that is, defining the situation in terms of perceptions and representations. One could argue that the less directed an interview is (where at most the interviewer says one thing only 'Start from wherever you wish') the better it is when trying to explore the ways in which an individual elaborates their personal history and gives meaning to their life (these are, indeed, called 'narrative' interviews). By contrast, an interview that focuses on themes seems to be the best when seeking to reconstruct specific experiences and relations; however this type of interviewer-led, focused questioning may create resistance and blocks in the narrator.
- Life stories or autobiographical accounts. This is a person's life story as told, through conversations and interviews, between themselves and an interviewer: a life history is a story about an individual, their experiences, strategies, vicissitudes and emotions. If the individual tells of events in which the have taken part, that is, if this personal account focuses on society at large and on social events, then it is termed oral history.

Among the indirect techniques are:

- *Written, requested/commissioned autobiographies*: these are autobiographies that subjects, who are considered to be of interest for research, are expressly asked to write. Here an autobiography is considered to be a written account of a person's whole life, a first person account written by the subject him/herself over a fairly limited period of time which encourages him/her to reconstruct the past. This account may be oriented, directed, through questions or through the provision of a list of the most useful themes/subjects. One problem that should not be ignored is how the subjects may react: any relationship with the institution that collects these autobiographies could influence the account. Furthermore, such autobiographies risk being affected by hindsight, by *a posteriori* rationalisations of past events.[18] The ideal type of autobiography is that which has been produced spontaneously – still a rare type to find in the social sciences today.
- *Biograms*: this is a particular kind of written autobiography, not only does the researcher ask for it but s/he also controls and checks it too. In other words, these autobiographies are fragments, or life events, briefly described by subjects in response to precise suggestions on the part of

the researcher. Like other sources listed above, these methods are tending to disappear as the importance of oral testimony increases.

- Both *personal and day-to-day diaries* have been used considerably in the past, especially during the 1920s and 1930s when the Chicago School dominated social research. The main characteristics of a diary are: it is strictly personal and events are written down, simultaneously, even as they unfold. These subjects become the archivists of their own day-to-day experiences; indeed, it is often worth telling them what the survey is seeking to achieve and asking them to record personal experiences related to that specific problem. While a great many diaries have been published as literature, very little social research has as yet made use of them: in fact they are not only rare but also it is difficult to generalise from them. Diaries are already widely used in psychotherapy but could also be efficaciously used in social research to keep track of the way in which difficult situations or crises are evolving (reactions to the news of serious illness, redundancy and the way in which responses change over time)[19] and, more generally, to discover how major, socially important events may affect individual behaviour, especially those events whose characteristics can be checked through other sources.
- *Letters*: research based on letters has become increasingly rare in sociology: as telephones have spread, so communicating by letter has become more and more unusual in today's society. However, letters often do still become a means of communicating during periods of enforced separation (wars or emigration). Letters were however used – as already described – in the famous study carried out by Thomas and Znaniecki (1918–20) on Polish emigrants to the United States.
- *Life history calendars or LHC* (also called biographical/life history matrices): biographical longitudinal type information can be collected on individual charts in the form of a matrix which specifies both the type of events the individual has experienced and when those events took place. The LHC format is usually a large grid (an example is given in Table 2.1). One dimension of the matrix is the behavioural patterns being investigated; the other dimension is divided into the time units for which these behavioural patterns are to be recorded: usually, the year, or years, in which such events took place and the age of the individual at the time are both recorded. In other words, these charts usually list the features and events which punctuate an individual's life (birth, end of education, marriage, first child, divorce …) horizontally (line-vectors) and the temporal meaning of the events vertically (row-vectors). A life history calendar can have two main advantages for collecting retrospective survey data. First, it can improve recall (and thus the quality of retrospective data) by increasing the respondent's ability, both visual

and mental to place different activities within the same time frame and to cross-check the timing of an event across several different domains. Second, very detailed sequences of events are easier to record within an LHC than with a conventional questionnaire (Freedman *et al.*, 1988).

Notes

1 For further details see Courgeau and Lelièvre (1988); Kasprzyk *et al.* (1989); Davies and Dale (1994).
2 The consequence of this advantage is that the dependence structure between the repeated observations must be identified, and this has become a delicate matter in the treatment of this data (Capursi, 1993) (see Chapter 4 for details).
3 For a discussion on the advantages and disadvantages of cross-sectional versus longitudinal studies data see, among others: Coleman (1981); Davies (1994); Dale and Davies (1994); Blossfeld and Rohwer (1995); Rajulton and Ravanera (2000).
4 The probability that a person will be selected twice to take part in a sample of 1,000 persons extracted at random from a total population of 20 million is 1 million out of 400,000 million, i.e. 1 in 20 million.
5 Waves correspond to the number of times a panel study is repeated.
6 See Appendix 2.
7 See Appendix 2 for further details.
8 The book was published in 1974 and is based on Elder's work with the Oakland cohort (167 people born in 1920–21). One chapter investigates the impact of World War II and includes the results from comparative studies with a younger birth cohort, the Berkeley Guidance Study (see paragraph 1.2 for details). Contrary to expectations at the time, Elder and his colleagues found that a great many of the children in the Oakland sample succeeded in rising above their childhood disadvantages and in achieving a full life to the seventh decade. The Oakland children encountered depression hardships after a relatively secure phase of early development in the 1920s, and they left home after the worst years of the 1930s for education, work and family. This historical pattern differed strikingly for the members of the Berkeley Guidance study born at the end of the 1920s. These children experienced the vulnerable years of childhood during the worst years of the Great Depression, a period of extraordinary stress and instability. Their adolescence coincided with the 'empty households of World War II' when parents worked from sunup to sundown in essential industry (Elder, 1974; Elder, Modell and Parke, 1993).
9 The IDA contains information from the following statistical registers held at Statistics Denmark: the Central Database on Salary Information (COR) administered by the Central Customs and Tax Administration; the Register of Population Statistics; the Educational Classification Module (UKM)/the Register of Education and Training Statistics; the Employment Classification Module (AKM); the Register of Income Statistics; the Register-based Statistics of Establishments and Employment (EBS); the Register-based Labour Force Statistics (RAS); the Register of Unemployment Statistics.
10 Cf. Ashenfelter and Solon (1982); Duncan and Kalton (1987); Duncan, Juster and Morgan (1987); Solon (1989); Duncan (1992); Rose (1994); Ghellini and Trivellato (1996); Trivellato (1999).
11 In Duncan's opinion (1992) a panel makes gathering long-term data both easier and more efficient: the shorter time period that elapses between waves together with the possibility this offers for comparing retrospective information with that gathered during the previous waves, ensure that the information gathered is high quality data.

12 Cf. Duncan (1984); Bane (1986); Muffels (1992); Muffels and Berghman (1992); Duncan *et al.* (1993); Gershuny (1998).

13 The period between two changes of state (for instance, from 'being employed' to 'being unemployed') is called an *episode*, *waiting time* or *spell*. The change from one spell to another is commonly termed a *transition* or (terminal) *event* (Taris, 2000: 95).

14 Life-course is simultaneously a relative, relational and dynamic concept.
- Above all, it is different from the concept of a life 'cycle' (basically a notion derived from biology) because no single phase of an individual's life should be read as a simple return to preceding phases, but ought always to be seen as a subsequent construction, that is, as the outcome of processes of accumulating and integrating experiences: emphasis is laid on the continuity of development and change within the individual's life span (Saraceno, 1986; Ongaro, 1995).
- Second, the phases it is made up of vary both in space and time and are of a social as well as a biological nature: thus they are influenced by both cultural and material differences and take on both different meanings and different values in different epochs and different societies.
- Third, even though the life-course approach has, traditionally, focused on the study of people, someone's life-course cannot be isolated from the situations in which s/he is immersed because individual choices take place within historical and geographical situations, i.e. in different 'macro' contexts. At the same time, people who may live in the same geographical or temporal context may experience different situations depending on the social relations they are involved in. Thus there are diverse mechanisms, social norms, which 'impose order and restrictions' on life-courses (Mayer, 1991 and 1996; Elder, 1992). Institutional/Public rules and regulations offer a normative context and a calendar/timetable of times within which a 'normal' life will, usually, develop, proceeding by means of continuous passages from one status to another. This normative context/calendar will tell an individual when is the 'right' time, or age, for getting married, for having children, for retiring, etc. With their ability to sanction actions, these institutions are also able to decide whether a life-course is 'regular' or 'irregular', i.e. 'normal' or 'not normal' and whether an individual does or does not have the right to services (Stone, 1991).
- Last, if society and social change can be studied through changes and evolution of the life-courses of the individuals who make up that society then the view obtained of socio-demographic dynamics is, essentially, longitudinal (Billari and Rosina, 1999). According to Giele and Elder (1998) the paradigm of a life-course is a fundamental element in the study of people's life-courses.

15 See Featherman (1980); Tuma and Hannah (1984); Sandefur and Tuma (1987); Mayer and Huinink (1990). Mayer (1991) identified four categories of mechanisms that 'impose order and limits' on the life history of an individual, all of which differ greatly both in themselves and between one individual and another:
1. institutional careers: the life-course of an individual is structured at the social level by the sequence of roles created in each of the phases of a life-course (e.g. through the school system);
2. public intervention and regulation: the public sector defines and standardises both the start and the end of many events;
3. cumulative contingencies: this is the after effects of restrictions created by the past which took place at an earlier point in the life-course of the individual (e.g. the effects of staying on at school on future work career and family life);
4. the overall conditions under which individuals belonging to diverse birth cohorts may end up (Billari, 1998).

16 Within the history of life events the unit analysed may not only be the individual him/herself, but also the event itself. In this case the line-vectors concern individual histories relating to that event (the birth of children, hospitalisation, retirement, etc.). Transitions may be considered instead (the transition from a part-time to a full-time job, from maternity to work), thus considering more than one data collection episode about events that are continuous over a period of time.

17 For more details see section on retrospective studies in Chapter 3.

18 As Gobo (1997: 39–40) reminds us, the process of remembering/re-evoking information is considered, by Cognitivists, to be a process of *construction* in which the person remembering will add something of their own to the event. After having carried out this type of 'incorporation' the individual is no longer able to distinguish between what s/he saw or heard and what they inferred (Loftus and Palmer, 1974). Subjects may even invent details that did not exist within the event being recalled because of *scripts* which encourage them to reconstruct the past in a stereotypical manner (Cantor and Mischel, 1977; Mandler and Johnson, 1977; Bower *et al.*, 1979). In other words, the content of any memory will be a mixture of the event which really did happen and other later additions.

19 Individual strategic autonomy is extremely important: a trajectory of need and/or crisis is characterised by the inextricable bond between the catalysing event and the individual's strategy which, by re-defining the event, allows the individual to adapt to it. Consequently the way in which a subject adapts to a problematic event is a process of construction within their life-course (Negri, 1990: 184–5).

3 The issues of data collection and comparability within longitudinal research

Some examples

This chapter aims to look in detail at some of the more important examples of longitudinal studies and, at the same time, examine the crucial issue of comparability in dynamic research. Currently, many independent national longitudinal studies are in operation in different countries of Europe and in North America. Although the contents of the questionnaires used may vary, in order to reflect the particular research purposes and policy interests of their sponsors, data are routinely collected (prospectively or retrospectively) on matters such as employment, family structure and changes in income, housing, consumption and health.

Prospective studies – an example of good practice: the British Household Panel Study (BHPS)

The BHPS was set up in 1991 by the ESRC Research Centre on Micro-Social Change[1] (now the Institute for Social and Economic Research (ISER) of the University of Essex). It is a high quality longitudinal file largely because it was designed only after the characteristics of existing HPSs had been carefully examined: in particular the Panel Study of Income Dynamics (PSID) (United States) and the German Socio-Economic Panel (GSOEP) (Germany) both of which inspired much of the BHPS. This denotes a willingness to try to increase the opportunities for comparing data, consequently, the BHPS could be considered to be emblematic.

The BHPS is based on a nationwide sample of about 5,500 households and 10,200 individuals (see Appendix 1 for details); data are gathered annually and the population is composed of all adult household members (16 years and over) who are resident in Great Britain (England, Scotland and Wales). The main aim of the survey is to study economic and social change, in Great Britain, at both the individual and the family level.

This British survey is of particular interest as the methods used to collect the information were very precise. Two aspects in particular should be highlighted:

1 the use of a pilot *micropanel*. This was made up of about 450 families and ran until 1994. This micropanel was organised to test versions of the questionnaires relating to each wave, also testing their longitudinal aspects;
2 the wealth of the information gathered. The data collected cover a vast range of themes which are important for the social sciences: family composition; income; participation in the labour market; living conditions; education; health; use of social services; division of responsibilities within the family; the economic strategies and choices of the family nucleus; and residential mobility. The questionnaire also asks both for information about any changes that may have taken place within the nucleus since the last annual interview and for retrospective information about the work, family and matrimonial histories of the subjects involved.

Example 3.1 The BHPS questionnaire package

The BHPS questionnaire package consists of:

- A household coversheet, which contains an interviewer call record, observations made by the interviewer about the type of family and type of accommodation and the final household outcomes. Cover sheets are produced containing the last known address of sample members. Moves discovered by interviewers during fieldwork are dealt with by interviewers, either by discovering a forwarding address or by creating a 'movers form' which is returned to the Institute.
- A household composition form which is completed, in most cases, at the interviewer's first contact with an adult member of the household. The interviewer gathers a complete listing of all household members together with some brief summary data of their sex, date of birth, marital and employment status and their relationship to the household reference person (HRP) – defined as the person legally or financially responsible for the accommodation, or the elder of two people equally responsible. Additional checks are required on the presence in the household of natural parents or spouses or partners, in order to unambiguously establish all relationships (for instance, secondary or 'hidden' couples).
- A short household questionnaire completed with the household reference person, which takes, on average, 10 minutes to complete. This contains questions about the accommodation and tenure and some household-level measures of consumption.

- The individual schedule takes approximately 40 minutes to complete and is applied to every adult member of the household (aged 16 or over). The individual questionnaire covers the following topics: neighbourhood, individual demographics, residential mobility, health and caring, current employment and earnings, employment changes over the past year, lifetime childbirth, marital and relationship history (wave two only), employment status history (wave two only), values and opinions, household finances and organisation.
- A self-completion questionnaire, which takes about five minutes to complete. Questions included are subjective or attitudinal questions particularly vulnerable to the influence of other people's presence during completion, or potentially sensitive questions requiring additional privacy. The self-completion questionnaire contains a reduced version of the General Health Questionnaire (GHQ) which was originally developed as a screening instrument for psychiatric illness, but is often used as an indicator of subjective well-being. It also contains attitudinal items and questions on social support.
- A proxy schedule is used to collect information about household members absent throughout the field period, or too old or infirm to complete the interview themselves. It is administered to another member of the household, with preference shown for the spouse or adult child. The questionnaire is a much shortened version of the individual questionnaire, collecting some demographic, health, and employment details, as well as a summary income measure.
- A telephone questionnaire, developed from the proxy schedule, for use by an experienced interviewer employed by the Institute. This is used when all other efforts to achieve a face-to-face interview have failed.

The questionnaires went through a series of major revisions, from the initial pre-testing through the two pilots, to produce the final versions used in waves one and two.

Web site: http://www.irc.essex.ac.uk/bhps/

The research group, made up of about 50 people, takes part both in drawing up and in the annual revisions made to the questionnaire: this encourages continuous collaboration between members who may have diverse technical and research skills.

It is a two-stage sample, with implicit stratification in the first stage units – postal areas – and, at the second stage, systematic extraction of addresses. Sampling was carried out using the Postcode Address File (PAF) for Great Britain (excluding Northern Ireland). This file lists postal addresses on a geographical basis and is the source most usually used for wide-ranging governmental surveys. In the first stage, 250 postal sectors were chosen using

systematic sampling of a stratified list of all PAF sectors: this gave the primary sampling unit. These primary sample units had already been, in their turn, stratified on the basis of the socio-demographic features of each postal sector, characteristics which were identified on the basis of 1981 census data. During the second stage, addresses were extracted from each sample unit using the same type of systematic procedure. Interviewers then got in touch with the family/families selected at the addresses chosen: if there were up to three families living at one address then all three would be interviewed; if, however, there were more than three families living at that address, the interviewers had to assign a number to each family and then choose three of the families at random.

Like the GSOEP, the BHPS uses a commercial research firm for the fieldwork, for coding and for the initial editing of the data collected. The Research Centre did intervene in these phases but only for training and overall supervision of the work done. The interviews for the BHPS are conducted by NOP Research, a commercial research organisation based in London, who carry out the fieldwork under contract to the Institute for Social and Economic Research (ISER). Until 1998, panel members were interviewed, face to face, on an annual basis. However, in September 1999, wave nine of the BHPS went into field using a Computer Assisted Personal Interviewing (CAPI) mode of data collection for the first time (Laurie, 2000).

It is not hard to obtain access to BHPS data: potential users have only to sign a form agreeing to respect the confidentiality of the data they obtain. The data are supplied free of charge, only the costs of any materials involved (photocopies, diskettes, etc.) have to be paid for. Furthermore, they are accompanied by invaluable, comprehensive documentation – a crucial element for a study of this magnitude as it allows users to evaluate the quality of the data available (Freed Taylor, 2000).[2]

A non-hierarchical user database is adopted for the purpose of analysis. This database is available both from Scientific Information Retrieval (SIR) and from Statistical Package for the Social Sciences (SPSS) (a software package for PC data management and analysis) and is accompanied by a detailed handbook. The user database is held at the University of Essex in the UK Data Archive, a centre dedicated to preserving and making available the main collections of sociological data currently in existence in Great Britain. It maintains a collection of over 4,000 significant social indicator datasets about all aspects of economic, political and social life. The huge quantity of datasets held by the Data Archive, together with the high quality of the data gathered, has allowed the University of Essex to become a 'Large Scale Facility' within the European Community's 'Training and Mobility of Researchers' programme and to obtain funding so as to be able to invite both senior and junior researchers to come and carry out research at the University itself.

Secondary analyses of BHPS data are much encouraged. Among the various initiatives set up to promote new approaches and analyses, and to offer suitable training and knowledge to the younger generation of researchers to help them to take up longitudinal research, are:

- literacy, or familiarisation, courses on BHPS data which are now run as part of the University of Essex Summer School;
- study grants (for research doctorates) which are being offered to both British nationals and foreign students as part of a project: Post-Graduate Research Opportunities in Economics, Sociology and Panel Data Analysis;
- periodical seminars which offer updates on research currently in progress;
- an annual meeting that is open to all users and offers the opportunity to discuss and resolve problems as well as presenting the results of work carried out and permitting an exchange of opinions and knowledge;
- a newsletter, distributed free of charge to all users.

Retrospective studies – how to develop a life-course study 'quantitatively': the German Life History Study (GLHS)

As stated above, the biographical approach often adopts non-standard data-gathering techniques (Fuchs, 1984; Voges, 1987; Olagnero and Saraceno, 1993). However, the GLHS is markedly different from such studies. The aim of this study is not merely to gather narrative sequences but is, rather, to collect data about life events and the more important activities of subjects (duration and frequency).

The GLHS is one of the few standardised studies of life-course carried out on large population samples (Brückner and Mayer, 1997). The main aim of the study is to reconstruct the social, historical and generational context within which the mechanisms that generate social inequality can be found. The passage towards the new socio-economic structure has spotlighted a series of previously unknown social risks and has profoundly influenced the way in which these risks are distributed throughout a life-course. Furthermore, individual and family life-courses differ considerably from one society to another and they are heavily influenced by any specific links between the state and the labour market. Thus it is becoming more and more important to develop a comparative perspective within life-course analyses.

The idea of developing a study like this first emerged during a project on intergenerational social mobility which was being carried out at the University of Mannheim and involved economists, sociologists and methodologists (Mayer, 1977; Mueller, 1978; Handl, 1988). As already described, the GLHS

has reconstructed the lives of men and women from different birth cohorts (seven in West Germany and five from East Germany (DDR)) using only one retrospective survey per cohort (Featherman, 1980).

The study was launched in 1979 and the first wave (carried out between 1981 and 1983) involved samples drawn from three birth cohorts (1929–31, 1939–41 and 1949–51) of West German residents (WGLHS). Other West German cohorts (1919–21, 1954–56 and 1959–61) were added later to gather data both about individuals who had had a different experience of war and about younger people. In the 1985–87 survey, the birth cohort of 1919–21 was added, and in 1988–89 data for the cohorts born 1954–56 and 1959–61 were collected. In 1997–98 a new wave on two birth cohorts, 1964 and 1971, was completed (employment biographies and labour market conditions) (Brückner and Mayer, 1997: 154).

After the fall of the Berlin Wall the study was extended to include the former East Germany (DDR) to study life-courses that had been affected by marked historical and social discontinuity. The cohorts included in the East German Life History Study (EGLHS) were those of 1929–31, 1939–41, 1951–53, 1959–61 and 1971 (Huinink *et al.*, 1995). Data on the first four of these cohorts (a total of 2,330 subjects) was collected in the period September 1991–October 1992; the survey group went back to the East German respondents with a written questionnaire in 1993 and interviewed them again in 1996–97, using both computer-aided telephone and face-to-face interviews, to cover the entire transformation process and explain its outcome in a life-course framework. In 1996–97, about 1,400 persons (about 61 per cent of the initial sample) were interviewed: the interviews focused on life histories since December 1989. And in the same year, for the first time, information was gathered about 600 women and men from the 1971 birth cohort: this cohort was chosen to monitor entry into the labour market, family formation and fertility behaviour under the extreme conditions of system transformation (Brückner and Mayer, 1997: 154; Solga, 1998). So far a total of 8,000 life histories have been put together which cover almost a century of German history.

The aim of the GLHS questionnaire is, above all, to depict the 'natural history' of individuals as accurately as possible and, at the same time, to render them quantifiable. The themes studied are: the characteristics of the family of origin and family history; education and professional training; residential history; work, income and consumption; social, religious and political participation; friendship and other informal networks; health and medical history.

One of the innovations introduced by this study is that systematic information is gathered about all members of the family nucleus. The questionnaire contains detailed questions about the family nucleus the individual is part

of at the time of the interview and about the interviewee's family of origin, (age, education, profession of father, mother, brothers/sisters, information about his/her parents' marital history and frequency of contacts with the family of origin).

A second, somewhat controversial, innovation is the way in which the interview is carried out, which has been changed over time. During the first two surveys only face-to-face interviews were used, creating a myriad of difficulties (long training for the interviewers and the costs of this training, wide geographical dispersion of the reference population) which eventually persuaded the organisers to start using telephone interviews adopting the Computer Assisted Telephone Interviewing (CATI) technique for interviews.

Example 3.2 The CATI technique

The CATI technique was first developed in the United States during the 1970s. It consists of a telephone interview during which the questionnaire can be seen on a computer: the text of the interview appears on-screen in front of the interviewer; replies are typed in immediately and memorised. The computer manages the course of the interview (for example, in the case of filter questions) and automatically highlights any incongruencies – this reduces the possibilities for interviewer error. This technique offers a series of advantages: random choice of telephone numbers for the sample; development of questions and contemporaneous codification and treatment of data; checks on interviewers; and rapid data elaboration. It also allows questions to be rotated; numbers called are recorded should further checks be required concerning answers given; checks can be made on the logical coherence of answers; and, lastly, the phone numbers of those who do not wish to be interviewed can be noted. Obviously this technique requires the use of either structured questionnaires or structured, and not too complex, interviews. For the drawbacks of this method see Frey (1989); Biorcio and Pagani (1997), Corbetta (1999).

This technique makes it easier and cheaper to train interviewers because it uses a group that is in one place and easy to reach for training, supervision, substitution and for checks. The telephone interview technique, however, does pose problems when collecting detailed retrospective information about life histories because it encourages the tendency to give stereotyped, superficial, or hasty answers, partly because there is less time available (Groves *et al.*, 1988; Herzog and Rodgers, 1988),[3] also the nature of the telephone itself means that questions have to be reduced to the bare essentials.

The time required for an interview varies considerably (there is a strong correlation with the age of the interviewee). Although, in some cases, interviews have been known to take as long as six hours, the average, for a face-to-face interview, is about 80 minutes for the cohorts born between

1929 and 1951 and somewhat over two hours for those born earlier, 1919–21. Like the BHPS, the GLHS uses external commercial research organisations to carry out interviews and to codify and input the data gathered (Brückner and Mayer, 1997).

These data are available to the public and are distributed by the Max Planck Institute of Human Development and Education in Berlin (Centre for Sociology and the Study of the Life-course), which has been managing the survey since 1983.

The issue of comparability within longitudinal research (European Community Household Panel, Panel Comparability Project, PSID-GSOEP equivalent data file, European Panel Analysis Group datasets, Consortium of Household Panels for European Socio-Economic Research project)

Comparison is a very important cognitive activity in both the human and the natural sciences: it is an essential ingredient for every cognitive activity, hence, for scientific knowledge too (Marradi, 1982, 1985). As Fideli (1999) complains, (see Fideli on the question in general) comparison is still too little used in the social sciences today. One reason for this is the lack of available data which is suitable for comparisons. In many countries very little data production prioritises, or even considers prioritising, making their data comparable with others' data. Data gathering, codification activities and the way in which sources are structured, are all strongly influenced by existing national conventions which may, in any case, change considerably over time. Consequently, data can rarely be compared with any real exactitude (Øyen, 1993; Hantrais, 1996; Hantrais and Mengen, 1996).

These problems are exacerbated in the field of longitudinal data collection because such data involve both variability over space and variability over time. Comparative studies and historiographic studies are inextricably linked: a historical perspective leads inevitably to a comparative study of society (Wright Mills, 1959).

Many attempts, both *ex ante* and *ex post*, have been made in recent years to improve the potential for comparison among existing HPSs.

Ex ante *attempts*

The realisation that there was a pressing need to have comparable longitudinal data available has inspired the setting up of an important survey at the European level, the European Community Household Panel (ECHP) or Europanel, which is running concurrently with individual national

longitudinal studies. The ECHP is planned for a total duration of nine years. It was launched in 1994 in the then 12 EU member states and is based on a probability sample of 60,819 households drawn from the EC member states in a proportion that reflects the size of each state's population (see Appendix 2 for further details). Since then, Austria and Finland have joined the project. From the fourth wave onwards, similar cross-sectional information extracted from administrative registers and the national Living Conditions survey will be available for Sweden.

Thus the ECHP offers a unique opportunity for future comparisons. Even though each member state organises its own data collection for itself, each dataset has a series of common characteristics that favour trans-national comparisons (Verma, 1997):

1 the national surveys are co-ordinated by Eurostat;
2 there is a common nucleus which serves as the starting point for developing the individual national questionnaires. The requirements of comparability do not necessarily imply that the same survey tools need be used in each country. Indeed, because of legal and institutional differences, some questions have to be formulated in different ways to obtain information that can later be used for comparisons (Marradi, 1982, 1985; Fideli, 1999);
3 common survey procedures (annual interviews and specific follow-up rules);
4 common standards for dealing with the data obtained (construction of variables, weighting criteria, recording and cleaning data, generation of derived variables, etc.);
5 common sampling procedures (size of sample, probability selection procedures, rules for finding subjects, etc.);
6 common paths of analysis developed by an international network of researchers.

Naturally, there are differences between nations as regards the response rate. Also, the degree of harmonisation between the various national questionnaires is still not considered to be entirely satisfactory.[4] Furthermore, it is still difficult to gain access to Europanel data as only the data on three member countries (the United Kingdom, Ireland and Portugal) can be freely distributed by Eurostat. Other countries, (Germany, Spain and France) restrict access to users with a specific contract. Any other national files can only be consulted at Eurostat itself (Eurostat, 1996a, 1996b) but Eurostat has prepared a public version of the survey, the Longitudinal Users' Database, a file of data which have been rendered anonymous, but these only relate to some of the original variables included in the survey. Any request to consult this file must come from an official organisation and access is only permitted

after payment of a sum which varies according to the category of each user (Marlier, 1999).

Example 3.3 The ECHP survey

Since the first ECHP results became available, there has been increasing demand, from both inside and outside the Commission, for ECHP-based statistics. Many researchers and other users have also expressed strong interest in having direct access to the data. However, ECHP micro-data contain information considered 'confidential' in terms of the EU Council regulation 322/97 of 17/2/97 on Community Statistics. Therefore, direct access to these original data has had to be more restrictive that would be desirable for a full exploitation of the data. In view of this, Eurostat decided to develop, together with National Data Collection Units (NDUs),[5] a set of rules allowing for easier direct access to 'anonymised' ECHP micro-data, without jeopardising both the necessary conditions of data confidentiality and the value of the data. In this context, in November 1997, Eurostat proposed that NDUs should create a user-friendly and widely documented Longitudinal Users' Database (UDB) that would meet various 'objective anonymisation criteria'. Here 'objective' is used to mean that once these criteria have been applied to the various ECHP files, there should be no risk that an individual statistical unit could be identified through 'all the means that might reasonably be used by a third party to identify the said statistical unit' (EU Council regulation 322/97 of 17/2/97 on Community Statistics) (European Community Household Panel Longitudinal Users' Database, Waves 1 and 2, Manual, 1998). In the UDB, the original variables have been fully reorganised, grouped together and standardised, so that they do not reflect the structure of the questionnaire any more, and analytical variables derived from original variables have been added. The names of the variables are the same in each wave.

Ex post *attempts*

Four innovative *ex post* initiatives to use HPS files have been set up with the aim of facilitating comparative research based on longitudinal data.

The first is co-ordinated by CEPS/INSTEAD[6] in Luxembourg: the Panel Comparability Project (PACO). This represents an innovative and centralised attempt to create a database of comparable variables integrating micro-data from various national household panels over a large number of years. This co-operative project was funded by the European Community. The funds have been used to set up a database of information drawn from seven European and non-European HPSs (Great Britain, Germany, France-Lorraine, Luxembourg, Poland, Hungary, United States). The database covers about 150 variables which have been rendered comparable through

a process of transformation which has created identical names, labels and methods and has, thus, been able to create a common data structure.

Example 3.4 Variables in the PACO file

The variables currently available can be grouped under the following headings: demography (16); income and financial situation of households (66); labour market (29); education (3); housing (1); use of time (4). The names of the variables are constant apart from time indicators (year); the first character indicates the level at which information is available: P = person, G = group (only Luxembourg and Lorraine), H = household. For example, the variable P84001 identifies the total of personal wages and incomes for the year 1984. Two files have been created for each geographical context: one with information at the level of the family and the other at that of the individual level (including children). As well as files relating to each year, there is also a file which contains information that does not change from year to year (sex, year of birth, educational qualifications, etc.). PACO also contains two longitudinal files, relating to Germany and the United Kingdom, which offer biographical information collected retrospectively.
Web site: http://www.ceps.lu/paco/pacopres.htm

This 'harmonised' file structure allows researchers to carry out transnational longitudinal comparative surveys more easily, offering excellent research opportunities (Schaber and Schmaus, 1996; Schmaus and Riebschlager, 1995). The importance of this harmonising project has turned CEPS/INSTEAD into a 'Large Scale Facility' financed by the European Union's Training and Mobility of Researchers Programme,[7] and is able to obtain funds enabling researchers from European Union countries to go to Luxembourg to consult these files.[8]

Another *ex post* project aimed at aiding comparative studies of HPS data is the PSID-GSOEP Equivalent Data File set up by the University of Syracuse, in New York State. It involves the PSID (United States) and the GSOEP (Germany) (Daly, 1994; Burkhauser, Butrica and Daly, 1995, 1999). The Equivalent Data File was developed because, although both these longitudinal surveys gather similar data about family composition, income, employment, housing conditions and demographic characteristics, the PSID and the GSOEP use different methods to collect their information. Consequently, it is difficult to directly compare the original two files.

The PSID-GSOEP equivalent database is, thus, the product of an attempt to render two sets of data more homogeneous. The first version covers the years 1984–89; the second continued with the work of standardisation up to 1994 and is made up of 11 waves. It is composed of two matrices (one relating to PSID data and the other to GSOEP data) with about 30 variables rendered homogeneous through being assigned the same names, labels and

formats. The sectors involved are: demographic information (8); employment (3); income (8); macro-economic indicators (1). To assure cross-national comparability, a number of variables in the PSID-GSOEP equivalent file had to be constructed/generated.

Example 3.5 Generated variables in the GSOEP

There is no direct report of annual work hours in the GSOEP (Butrica, 1996b). This variable was constructed using information on employment status in the survey year, average number of hours worked per week, and the number of months worked in the previous year (reported in the activity calendar). The most complex of the generated variables is the GSOEP measure of total household income after taxes and transfers (post-government income). In the PSID, the construction of this variable is fairly straightforward, since the data are already available in a yearly frame and a tax estimate is provided. However, constructing a comparable variable in the GSOEP was a much more complicated task. To create this variable, all monthly income amounts had to be annualised. Next, annual tax burdens for all households in the GSOEP had to be estimated by using a tax simulation package that modelled the German tax system (for a fuller discussion of the creation of these variables see Burkhauser *et al.*, 1995; Schwarze, 1995; Butrica, 1996b).

It is important to remember that what was, formerly, the PSID-GSOEP Equivalent File is now being substituted by the Cross-National Equivalent Files (CNEF), that contain harmonised panel data from Canada, Germany, the United States and the United Kingdom. The newest release includes data from the PSID for the years 1980 to 1997, data from the GSOEP for the years 1984 to 1997, data from the Canadian Survey of Labour and Income Dynamics (SLID) for the years 1992 to 1994 and data from the BHPS for the period 1991–98.

The European Panel Analysis Group (EPAG) project was launched in 1990 under the direction of the Institute for Social Research at the University of Essex.[9] This research group aims to monitor the development of HPSs in the European Union and, at the same time, to contribute to and encourage the spread of comparative longitudinal research in the field of family dynamics, work, poverty and social marginalisation

The EPAG is a consortium of European social and economic researchers who have been collaborating since 1990 in the development and analysis of HPSs within the European Union. Most recently it has been engaged in the study of flexible labour and its impact on earnings and poverty under a Eurostat contract, and in a programme of research on social exclusion as part of the European Union's Targeted Socio-Economic Research programme. The group has set up new comparative datasets based on five-year sequences of the British, German and Dutch national household panels,

and is analysing the early data from the ECHP. To date, most of the research has been in the fields of family formation, employment, household income and deprivation. The EPAG dataset can be accessed through the European Centre for Analysis in the Social Sciences (ECASS) programme – which is another Large Scale Facility for the Social Sciences that offers access to files held in the Data Archives of the University of Essex.[10]

Finally, we should also mention the Consortium of Household Panels for European Socio-Economic Research (CHER), whose aim is to develop and enhance a comparative database for longitudinal household studies by harmonising and integrating micro datasets from a large variety of independent national panels and from the ECHP. In order to promote the accessibility of both comparable and longitudinal micro data, the Consortium will create an international comparative micro database CHER/PACO containing longitudinal datasets from many national household panels and from the ECHP which will be complemented by key information from existing macro/institutional datasets that are linked to the comparative database and supported by utilities for panel analyses.

This project will build on the work already carried out by the various partners of the consortium: Belgium, France, Germany, Greece, Hungary, Italy, Luxembourg, The Netherlands, Poland, Spain and Switzerland.

First, the consortium will create, update and/or integrate two existing international comparative databases:

- PACO Database;
- PSID-GSOEP Equivalent Data File.

Second, the CHER database will integrate longitudinal datasets in Europe over a much larger number of years from as many country household panels as possible and from the available country datasets present in the ECHP. Third, the database will be supplemented with data from both the United States and Canada. The final CHER database will contain comparable variables that have been transformed according to a common plan and will be built by using standardised international classifications where they are available. Information in these files will be available: (a) for households and individuals on the micro level; (b) for single years; and (c) as longitudinal information, all of them linked to meso and macro data.

The comparative database – held in a relational database structure where data are stored as system files for the statistical packages SPSS, SAS and STATA – will contain harmonised and consistent variables and identical data structures for each country included: identical variable names, labels, values and data structures. Each country's file will be adequately anonymised and can therefore be rated as a scientific use file. The database will be available on a CD-ROM and will be distributed to the scientific community, under appropriate rules for confidentiality and data protection. For the data coming

directly out of the ECHP the Consortium will adhere to the rules set by Eurostat.

The Consortium will also set up three small databases containing key information about: (a) macroeconomic information and social information; (b) social security; (c) employment policies. These three databases will have a link to the relevant variables in the CHER/PACO micro database. They will help in the interpretation of results from national and cross-national research with both the comparative CHER micro databases and original datasets from the panel studies. The macroeconomic information, social information database will contain key information about demography, labour force participation, unemployment, social protection, labour costs, price indices, purchasing power parities and similar items. The information will be extracted from existing publications/databases such as Eurostat-CD (yearbook), New Cronos, ESSPROS, OECD series and some already existing comparative welfare state datasets. The data for the social security database will be extracted from Mutual Information System on Social Security (MISSOC) publications and the data for the Employment Policies database from Mutual Information System on Employment Policies (MISEP/ERSEP) publications (http://www.kub.nl/~fsw_2/asz/tisser/research/Cher.htm).

Notes

1 The *Research Centre* was specifically set up in order to study social change at the micro level.
2 As will be seen in Chapter 4, the problem of the quality of data is particularly pressing when dealing with longitudinal studies.
3 For more information about telephone interviews see Lavrakas (1987) and Frey (1989).
4 For further information regarding the advantages and disadvantages of the Europanel see Ditch *et al*. (1998: 2–3).
5 NDUs are responsible for selecting the national sample, adapting the questionnaire to national standards, and carrying out the fieldwork, basic data processing and editing at the national level. The 14 NDUs are:
   - Austria: The Interdisciplinary Centre for Comparative Research in the Social Sciences (ICCR), IFES/FESSEL;
   - Belgium: UIA, UFSIA (Centre for Social Policy), University of Antwerp;
   - Denmark: Danish National Institute of Social Research;
   - Finland: Statistics Finland;
   - France: Institut National de la Statistique et des Études Économiques (INSEE);
   - Germany: Statistisches Bundesamt (StBA), Statistical Office of the Länder;
   - Greece: National Statistical Service of Greece (NSSG);
   - Ireland: Economic and Social Research Institute (ESRI);
   - Italy: Istituto Nazionale di Statistica (ISTAT);
   - Luxembourg: Centre d'Études de Populations, de Pauvreté et de Politiques Socio-Économiques (CEPS);
   - The Netherlands: Centraal Bureau Voor de Statistiek (CBS);
   - Portugal: Instituto Nacional de Estatistica (INE);
   - Spain: Instituto Nacional de Estadistica (INE);

- United Kingdom: Social and Community Planning Research (SCPR); in 1997 the Institute for Social Research (ISER) was designated the NDU for the British component of the ECHP. British data for Wave 4 (1997) of the ECHP consisted of re-coded derived variables. From Wave 7 (1997) of the BHPS 1,000 low income households selected from the ECHP sample have been added to the BHPS. This ECHP subsample has been merged with the BHPS sample and is interviewed annually.

6 Centre d'Études de Populations, de Pauvreté et de Politiques Socio-Économiques – International Networks for Studies in Technology, Environment, Alternatives, Development.

7 In March 1995, the Study Panel on Social Sciences, set up by the Directorate General XII for Science, Research and Development, identified CEPS/INSTEAD as one of the four European installations which might qualify as large-scale facilities for training and research in the social sciences. In November 1997, the Directorate General XII of the European Commission selected CEPS/INSTEAD for support under the Training and Mobility of Researchers (TMR) programme ('Access to Large-Scale Facilities') during a first period of two years (1 April 1998 – 31 March 2000).

8 Between 1997 and 1999 the PACO Project organised a training course that was supported by the Training and Mobility of Researchers (TMR) Programme of the European Community. The aim of the Panel Comparability training workshops was to disseminate the knowledge required for informed use of PACO and to enable social scientists to use cross-national and truly comparative panel data on a regular basis.

9 EPAG is made up of members drawn from the following research institutes/bodies: Institute for Social and Economic Research (ISER); German Institute for Economic Research (DIW), Berlin; Economic and Social Research Institute (ESRI), Dublin; Centre for Labour Market and Social Research (CLS), Aarhus; Tilburg Institute for Social Security Research (TISSER) and the Work and Organisation Research Centre (WORC) of the University of Tilburg and the Department of Sociology and Social Research of the University of Milano-Bicocca.

10 The already cited UK Data Archive and the Economic and Social Research Council Qualitative Data Archival Resource Centre (Qualidata). Qualidata was established in 1994 to ensure the long-term preservation and availability of a wide range of qualitative material. The Centre's remit includes providing an information resource on the location and accessibility of qualitative research material in general, as well as advice and training on the secondary uses and re-analysis of qualitative data. Within the Centre is the Peter Townsend National Social Policy and Social Change Archive.

4 Some problems connected with longitudinal research

We will now look at some aspects concerning the quality of longitudinal data. Even though, as was argued in Chapter 1, dynamic data offer a highly innovative and precious tool for the analysis of social phenomena, they do, nonetheless, have certain inherent disadvantages.

Discussion about the advantages and disadvantages of longitudinal research really began in the 1960s. Two important studies are worth citing here. The first is by Rene Zazzo (1967) of the University of Paris and the other (Wall and Williams, 1970) is the Report of the National Foundation for Educational Research that was commissioned in 1967 by the UK Social Science Research Council to identify the distinctive contribution longitudinal studies could make to the development of the social sciences. As Rajulton and Ravanera (2000) reported, after careful evaluation of the advantages and disadvantages of dynamic studies, neither of them was particularly in favour of pursuing longitudinal studies. But things have changed since that report was produced. Nowadays – despite the practical difficulties and the complexity of the structure of longitudinal studies – easier access to longitudinal data enables researchers to undertake empirical analyses of social change that were not possible before (Hallinan, 1997).

The limitations of repeated cross-sectional design

The obvious advantages of cross-sectional designs over longitudinal ones (e.g. saved time, less expense and absence of attrition) are compelling when the research question does not involve continuity and change (Copeland and White, 1991: 20).

However, because they do not use the same sample, trend studies only enable change to be analysed at the macro level (e.g. comparisons of the proportion of the population that is below the poverty line at time t and at time t–1). Given that the same individuals are not followed over a period of time, i.e. subjects are not re-interviewed, such studies are not suitable if one

is seeking to identify the causal mechanisms that govern social change (Menard, 1991). Furthermore, it is very hard to distinguish between the effects of age and cohort:[1] the principal limitations of repeated cross-sectional design are indeed its inappropriateness for studying developmental patterns within cohorts and its inability to resolve issues of causal order. Thus, cross-sectional type data cannot be a suitable source of information for identifying changes in behaviour that are the result of growing older/ageing (Saraceno, 1986).

Finally, as has already been seen (Chapter 1), in the case of analyses of the processes of mobility/inertia within social phenomena (e.g. poverty) historical series of stocks or of net changes which can be identified from repeated cross-sectional surveys are of limited use and may even be misleading: what is required is a description of flows (gross changes), which are essential for any study of mobility between states (Duncan, 1992; Ghellini and Trivellato, 1996; Rose, 2000). Consequently, it should come as no surprise that conclusions, drawn on the basis of cross-sectional data, have often been challenged by analyses based on longitudinal data (Lieberson, 1985; Mastrovita, 1998).

Thus, more data are required to describe, empirically, the dynamic process which lies behind the cross-sectional snapshot (Davies, 1994). The gradual adoption of survey designs with, to a greater or lesser extent, marked longitudinal features, is the result of the requirement to find satisfactory answers to questions concerning the dynamics and the determinants of individual behaviour: questions that remain unanswered, or receive only partial answers, from repeated cross-sectional studies (Ghellini and Trivellato, 1996).

Problems connected with panel design

Prospective studies have unmistakable methodological strengths, but they are expensive and time-consuming (van der Kamp and Bijleveld, 1998). Both Duncan (1989) and Blossfeld and Rohwer (1995) have summed up the problems posed by panel studies as follows:[2]

- Above all, physiological changes in the size of the sample (attrition) at each successive period of data-gathering, at each wave, represent a process of selective reduction in the number of subjects involved. Attrition occurs when respondents leave the panel – because of refusal to answer, physical incapacity of the respondent to provide information and/or failure to follow-up sample cases – after having participated in one or more consecutive waves, including the first wave of the study. These respondents are not contacted for later waves. Thus, attrition is cumulative: once a participant has missed one of the waves, they are lost for the remainder of the study (Taris, 2000: 20). This thinning process is not random: if those who leave the study are not typical of those who started it, any longitudinal data will become biased to the same extent

Table 4.1 Attrition rates in ECHP (%)

Country	*Attrition rate Wave 1 to Wave 2*	*Attrition rate Wave 2 to Wave 3*	*Attrition rate Wave 1 to Wave 3*
Belgium	10.0	n.a.	n.a.
Denmark	11.5	15.0	n.a.
France	11.0	n.a.	n.a.
Germany	8.0	n.a.	n.a.
Greece	9.0	9.0	17.0
Ireland	14.0	16.0	28.0
Italy	5.0	n.a.	n.a.
Luxembourg	6.0	n.a.	n.a.
The Netherlands	9.0	7.0	15.0
Portugal	4.0	n.a.	n.a.
Spain	12.0	9.0	n.a.
UK*	23.0	18.0	37.0
Austria		6.0	n.a.

Source: Eurostat, 1997.

Note:

* The exceptionally high figures for the UK is because households with the household interview completed, but with one or more uncompleted personal interviews within the households, were not followed-up by the UK NDU.

and this will produce a non-representative sample (a 'biased' sample). Attrition happens because of a refusal to continue, or through death and emigration, thus it can distort conclusions drawn on the basis of information supplied by that section of the sample that has survived/remained.[3] Table 4.1 shows the rate of attrition during the first three waves of the European Community Household Panel (ECHP).

- The problem of dealing with missing answers from a longitudinal point of view. Getting rid of the information from missing cases in cross-sectional studies is not a major problem but, if it is done at each wave of panel studies, may lead to severe distortion within the panel which will – at the end of this process – exhibit very different features from those it started out with.
- The fact there is a higher risk of measurement error than in cross-sectional data because errors accumulate over time (Fuller, 1987). For example, if data about income gathered at time *t* has errors, this could lead to false transitions appearing concerning phenomena such as poverty or unemployment (Duncan, 1992).
- The disentanglement of 'apparent' and 'true' change: this is one of the most complicated issues in panel analysis (Hagenaars, 1990: 18–19). It is usually assumed that the observed changes indicate true changes and are not just reflections of inaccuracies, of unreliability, in measurements. However, it is possible that some changes are due to measurement errors,

e.g. the misclassification of a unit at a given time into the wrong category of a discrete variable (Skinner, 2000). Measurement errors are conceived in terms of the difference between the measured value – that is, the value recorded in the data file – and the true value of a variable (Lessler and Kalsbeek, 1992: 370). Some of the changes observed may also have occurred because the questions are ambiguous, because respondents make mistakes when answering the questions, because interviewers and coders make errors during the data recording and processing phases. Consequently, it is very important to know how data have been collected.

- The fact that the nature of the answers given can be influenced by repeated participation in the panel (*panel conditioning* or *problem of sensitisation or time-in-sample bias*).[4] The problem of panel conditioning is 'the situation when repeated questioning of panel members affects their survey responses, either by altering the behaviour reported or by changing the quality of the reponses given' (Kalton *et al.*, 1989: 249–50). Precisely because they are repeated, panel studies tend to influence the phenomena that they are hoping to observe. During subsequent waves, interviewees often answer differently from how they answered at the first wave: this may either be because they have lost some of their inhibitions, or because they have acquired new information in the meantime, or because they have had new, different experiences during the time that has elapsed between one wave and the next. Subjects may also react differently during a second survey simply because they have had the experience of the first one. Thus, conversely, participation in a panel survey may also improve the quality of the data (Duncan, 2000).[5] The effect of sensitisation is particularly clear in longitudinal studies on electoral behaviour: Traugott and Katosh (1979) demonstrated how members of this type of panel often, over time, become more interested in political matters and participate to a greater extent in elections. As van der Kamp and Bijleveld (1998: 34) stated, some threats to internal validity (the plausibility of the observed relationship among presumed antecedent and consequent variables)[6] in longitudinal studies are: 1) *History*: historical events may modify the observations of a group of subjects. 2) *Maturation*: irrespective of the historical events taking place, subjects may show differential maturational changes in the course of a longitudinal study. 3) *Testing*: the measurement process itself may make respondents more 'expert' because of the practise, skill training, offered by repeated testing (repeated language proficiency tests generally increase students' abilities and their skill in completing such tests). To check for a sensitisation effect, panel studies may require control groups matched to the panel groups, further increasing the, already high, costs of this design. However, Sudman and Ferber (1979) concluded that, after the

initial wave, general purpose panels are unlikely to be distorted by 'conditioning' effects.

- Panel data offer information that is only related to pre-determined points in time (data are usually gathered annually, that is, at discrete time points). Thus the researcher cannot know about the course and evolution of events in the period that has elapsed between one collection time and the next. Furthermore, prospective studies are often limited to a few waves only and, consequently, cover only a short period of time. Indeed, there may well be a particular situation under way at the moment in time when information is being gathered which can distort individuals' answers (*fallacy of historical period*).
- Precisely because they offer information that is only related to specific, pre-determined points in time, panel studies are not able to reveal the time factor in historical processes. Both independent and control variables may change over time. Individuals are continually subject to changes in their personal status, they constantly acquire new work experiences, change their social relations and are perennially exposed to the effects of political, social and cultural changes. Rose argued that panel studies do allow checks to be made on the specific effects of age, period and cohort.[7] The frequent intervals with which the data are collected, and the multiple cohort aspects of the design, mean that cohort effects, period effects and developmental age effects, can all be monitored within the time span of the study (Bynner, 1996). However, Blossfeld and Rohwer (1995) did not agree. They maintained that these three effects are mixed together in panel studies, something which can only be corrected by adopting statistical procedures which are then used for many waves.[8]
- Panel studies often study the members of a specific cohort (see, e.g. the British National Child Study). In other words, they study individuals who were born, grew up and live in one, specific, historical period. This can be a danger if the researcher wishes to use this data as the basis on which to formulate general principles concerning life-courses (*fallacy of cohort centrism*).
- There must always be a time interval before a cause generates an effect, but in some cases this effect is generated almost instantaneously, such that the cause and effect appear to take place at the same point in time (Kelly and McGrath, 1988). Other effects may take longer to be generated. In this case there is a delay, or time lag, between cause and effect which must be specified in causal analyses. Two restrictive assumptions on which many prospective studies are developed are: that a cause will produce an instant effect, and that the real causal interval is approximately the same as the time interval that passes between one observation and the next.

- As well as the problem of the length of the time interval between cause and effect, one should also take into account the different forms the development of the time effect itself may take. While the problem of *time-lag* has often been discussed in the social sciences there is still very little information about the temporal shapes of effects, that is, how such effects develop over time (Kelly and McGrath, 1988). Effects may either appear in monotonic or linear forms, or they may be cyclical, or they may be even more complex. Consequently, the strength of the observed effect will depend directly on the timing of the panel waves, that is, on whether the panel places measurement points at a peak or at an ebb in the curve that expresses the temporal shape of how a change in a variable x, occurring at time t, effects the change in a variable y.
- When there is reciprocal causality other problems may emerge if the temporal structure of the effects of $x1$ on $x2$ and of $x2$ on $x3$ is different from the time intervals that already exist between cause and effect and from the way in which this effect develops. In such situations a prospective study is no use for researchers who want to identify the development of such time-related recursive relationships.
- A period of time needs to occur before an analysis of social change is feasible: a consistent number of waves is necessary to allow in-depth, long-term analyses to be carried out. Unfortunately, many national panels are not 'old' enough to permit dynamic analyses of social phenomena (see Appendix 2 for details).
- Lastly, the sheer complexity of such analyses, along with the fact that techniques of analysis that can be used with longitudinal data (see Chapter 5) have yet to be developed, should also be mentioned.

There are also problems that are inherent in the structure of the panels themselves.

First, panel data files are usually extremely large; the majority of existing household panels have initial samples of around 5,000 households and of over 10,000 individuals.

The high level of complexity of the structure of Household Panel Studies (HPSs) is also a problem. Such studies have two temporal dimensions (cross-sectional and longitudinal) and, furthermore, the data are usually gathered and stored at three levels: the household, the individual and the period/length of time (*spell-files*), where the unit analysed is neither the family nor the individual but the event (for further details see Chapter 2 at pp. 42–7). Thus, the structure of such data makes it possible to combine two separate units of analysis (family and individual) and, consequently, to create longitudinal files (by linking one wave to another through the use of original/unique individual and household identifiers or key variables) on the basis of

either prospective or retrospective longitudinal information which has been gathered at either the aggregate or the individual level.

Thus, HPSs are 'complex' in the sense that they consist of a number of different data structures or files, with differing focuses (some referring to the particular households studied at particular waves, some referring to individuals, some referring to particular events that the individuals surveyed have experienced) and often of repeated files, that have the same structures but relate to different points in historical time (that is, files describing respondents' circumstances in successive years/waves). This implies that analysts must apply some additional concepts to those included in the analysis of more straightforward survey datasets. The real value of this sort of dataset comes from the investigator's ability to link the various files together, connecting information in a number of straightforward ways. For example, attaching household-level information to the individual respondents, or connecting individual respondents' information over time. The crucial concept is that of a 'key variable' which identifies particular records within files as belonging to particular households or individuals. It is these key variables that tell us which parts of which files can usefully be joined together.

Example 4.1 Linking operations between HPS data files

In many cases, longitudinal data are released in the form of cross-sectional files pertaining to each set of interviews, and individual analysts may be required to link the files across time themselves before they can be used longitudinally. There are, in principle, three different sorts of linking or joining operations that can be done between HPS data files:

- Information organised at a particular level may be matched with other information organised at a similar level (e.g. evidence about someone in 1999 could be matched with information about the same person in 2000).
- Evidence organised at a particular level may be aggregated to a higher level of organisation (e.g. a file organised to provide information about every separate employment spell experienced by each respondent during the last year, might be reorganised to provide information at the level of the single respondent, e.g. about the number of changes of employment status during another year).
- Evidence from records organised at a higher level, may be distributed across records in files organised at a lower level (as when household-level information – e.g. concerning satisfaction with home/neighbourhood – is attached to all of the individual-level records of the members of each household).

Even when these linkages are provided in some form by the data collectors, however, analysts must often still decide how to carry forward information about household structure and family relationships from one interview to the next. In

other words, the analysts themselves must choose the basic unit of analysis around which the longitudinal file is to be organised. Deciding these issues will also help to determine what files are needed, that is, what sample(s) to include in the analysis (e.g. all individuals over the age of 16 who took part in a survey). Unfortunately, even relatively straightforward terms such as 'household' or 'family' lose a great deal of their precision when they are considered longitudinally. Families may undergo a substantial amount of change over the course of a year because households are dynamic, incessantly splitting and reforming in unpredictable ways. The impact of divorce, out of wedlock births, deaths, remarriages all change family composition and result in new families. Linking these new families with their predecessors in a way that facilitates our understanding of the impact of these transitions can be very difficult, since families can combine and recombine in many different ways over a given observation period. As a result, longitudinal linkages at the person level are generally the most satisfactory solution to the problem of linking records across time, since the person is the only unit of observation that is reasonably constant over time (Ruggles, 1990: 128, 1991; Walker and Ashworth, 1994: 32).

There are many different sorts of computer software available to carry out these linking operations. As well as special-purpose 'database management software' (e.g. SIR – Scientific Information Retrieval), this set of operations can also be carried out using some of the standard statistical packages (SPSS and SAS).

Web site: http://www.irc.essex.ac.uk/bhps/

Successful navigation through individual wave and cross-wave files is therefore a complex task and requires careful documentation. The user documentation is crucial to make longitudinal analysis both easier and more straightforward. It should contain essential information required for the analysis of data and information as well as information which will assist users when linking and aggregating data across waves (Freed Taylor, 2000). Bailar (1984) and David (1991) have identified a number of topics as being essential for the documentation of a panel dataset: design of the survey (sample, questionnaire, field procedures; coding, editing, linkage, treatment of missing data); design of the panel dataset (following rules, verification of linkage, periodicity);[9] facts of the survey (what is known about the data, including inconsistencies and anomalies); facts about the panel (information which is needed to understand how to condition data collected in later waves on data collected in prior waves); and, lastly, analyses (which record the completed work already carried out on the data).

An overview (Freed Taylor, 2000: 157–8) of the types of information which most panel dataset producers currently offer in survey documentation shows that information elements are: introduction and research study descriptions; statement on confidentiality and ethics statement; survey design

information (overview of the survey, questionnaires, sample design); survey context information (fieldwork details); advice on usage and data linkage (indication of analysis potential); sampling error, weighting and imputation information (procedures/algorithms used); data processing and coding information (procedures/techniques used); publication and analysis details (references to all publications relevant to, or based upon, analysis of the data); descriptive information (notes on terms/concepts used, technical information on database structure, notes on derived variable construction).

In the following parapgraphs we will look more closely at the complexity of two household panel studies. As a firstexample, at each wave a German Socio-Economic Panel (GSOEP) file will include:

- *at the cross-sectional level:* two files which contain information about all the members of the households sampled (size, area of residence, sex of household members, year of birth, relationship/rapport with the reference person, number of contacts, reasons for possible refusal to participate, interview method adopted) gathered by the interviewer but not directly gathered by the survey (available at both the family and the individual level);
- six files at the individual level which contain: 1) information drawn from the personal questionnaire (German sample of former West Germany); 2) information about the sample of (former) East Germans; 3) information about foreigners (Turks, Italians, Greeks, persons from ex-Yugoslavia and Spaniards); 4) information about children; 5) information about individuals who have temporarily left the survey; 6) information gathered through retrospective questions and contained in calendars (duration data). The original (1984) GSOEP sample was made up of two separate subsamples, each of which had its own sampling plan (see Table 4.2). The West German probability sample is representative of Germans living in the German Federal Republic (including former west Berlin) who were living in nuclear families in 1984. The sample was extracted on the basis of random selection at 548 sampling points. Within each sampling unit each interviewer received one address to start with and then chose a further 84 consecutive addresses of nuclear families using a systematic procedure to extract one family in seven. Of the 12 addresses that were left, only 10 were used for interviews: the other two served as reserve addresses;
- two files at the household level containing: 1) information gathered about the household drawn from the questionnaire; 2) duration data gathered at the household level;
- two different files of variables generated *ad hoc* (available at both the household and the individual level);
- *at the longitudinal level:* two files (individual and household) containing longitudinal information about all the members of all the households

Table 4.2 Sampling plan of the GSOEP

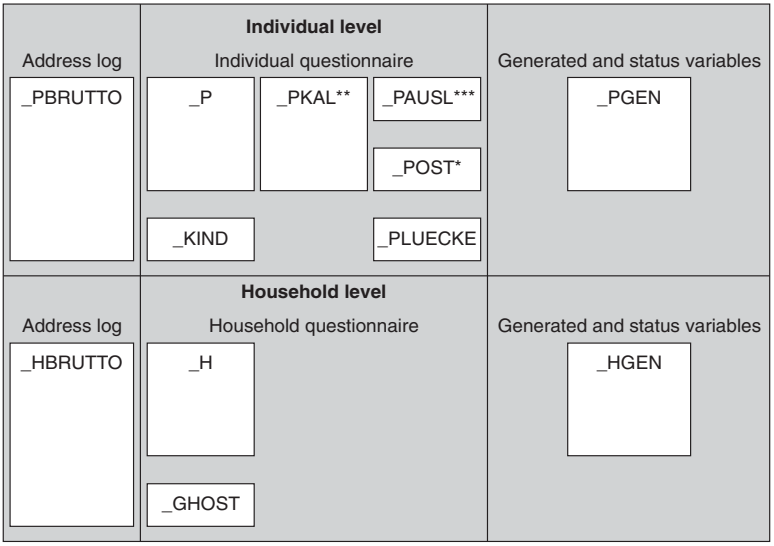
A 'West-German' residents (started in 1984)
• *n*=4,528 or 4,298 households*
• Head of household is either German or of another nationality than those in Sample B.
B 'Foreigners' (started in 1984)
• *n*=1,393 or 1,326 households*
• Head of household is either Turkish, Italian, Spanish, Greek or Yugoslavian.
C 'East-Germans' (started in 1990)
• *n*=2,179 or 2,071 households*
• Head of household at the time of the survey was a citizen of the GDR.
*D 'Immigrants'** (started in 1994/95 in two different subsamples)*
• 1994: Subsample D1 with *n*=236 households
• 1995: Subsample D2 with *n*=295 households
• At least one household member has moved from abroad to Germany after 1984.
E 'Refreshment sample' (started in 1998)
• *n*=1,000 households
• Random sample covering all existing subsamples A, B, C, D

Source: Frick, 1998 (GSOEP documentation, distributed to users on CD-ROM).

Notes:
* The first number relates to the full 100% version, the second relates to the 95% public use version of the GSOEP data.
** This sample has not yet been included in the 95% public use version.

who had been contacted at least once, including people/children who had never given an interview;
- two files containing weighting variables (which can be applied both to individual and household files);
- three files which store duration data: information gathered through the use of calendars;
- one file containing information about subjects who have left the survey (non-response).

To overcome the difficulties of dealing with the complex data structure of the GSOEP, a service is now being offered to users. Among other things, this includes a training course at the German Institute for Economic Research (DIW), published (tutorial) material, a user's handbook with regular updates and a Panel Newsletter,[10] which gives information about the latest developments. There is also a detailed list of available literature (also offered on floppy disk) which provides an overview of research results that used GSOEP data which have been published to date. An index supplies information about the contents of the research done with the panel data.



_ : Wave specification (A, B, C,...); * Waves G and H only; ** Waves A to G only; *** Waves A to L only

Legend
The file _PBRUTTO is the address log for all individuals in GSOEP households (i.e. respondents, children under age 16, and non-respondents).
All information from the individual-level questionnaire is located in the _P file.
_PKAL contains individual-level calendar information on income and work for 1984 to 1989. After 1989 this information is located in the _P file.
_PAUSL contains information on individuals in Sample B (Foreigners).
Information on Sample C individuals (Eastern States of Germany) is included in the _POST file.
Data on children under the age of 16 are found in the _KIND file.
Information on individuals who drop out of the survey but later return (temporary drop-outs) is located in the _PLUECKE file.
Generated and status variables for individuals are contained in the file _PGEN.
_HBRUTTO is the address log for all households in the GSOEP.
The file _H contains information from the household questionnaire.
Information from the 1990 questionnaire for households in Sample C (Eastern States of Germany) is located in the GHOST file. After 1990, information on Sample C households is included in the _H file.
Generated and status variables for households are contained in the file _HGEN.

Figure 4.1 The GSOEP data structure: yearly cross-sectional data

Source: Frick, 1998.

The structure of the Panel Study of Income Dynamics (PSID) data file is equally complex.

Before 1990, PSID main files for each interviewing wave consisted of a Cross-Year Family-Individual Response File, a Cross-Year Family-Individual Non-response File, and a Cross-Year Family File. Both the Cross-Year Family-Individual Response and the Non-response files had an identical file structure: one contained records for all individuals who were members of PSID family units interviewed in the most recent interviewing wave, while the other contained information on all individuals who were members of families interviewed in the past but who were not included in the most recent wave. The Cross-Year Family-Individual File stored both individual-level variables and family-level variables collected in the most current wave and in past waves. The Cross-Year Family File contained only family-level variables. Beginning with the 1990 data, the record format of the cross-year files exceeded the maximum allowed on most computing systems, and, consequently, a new file structure for the PSID data was developed. This new file format consists of separate, single-year files with family-level data collected in each wave (i.e. 23 family files for data collected from 1968 through to 1990), and one cross-year individual file with individual-level data collected from 1968 to the most recent interviewing wave. In this new scheme, each family file contains one record for each family interviewed in the specified year. The records in each file are identified by the Family ID for that year, are sorted by that variable, and contain the family-level variables collected in that year. The Cross-Year Individual File contains one record for each person who had ever been in a PSID family up to and including the current year. The records in the Cross-Year Individual File are identified by 1968 Family ID and Person Number and are sorted by these variables. The file also contains the Family ID of the family with which the person was associated in each year. With the new file structure, a moderate amount of data management is required to merge the family files with the individual file so as to create a traditional PSID cross-year family-individual file. The advantage of this new file format is that the files require the minimum amount of storage space. Since each file is considerably smaller than the traditional cross-year family-individual file, the PSID data in this new file format make less demands on computing resources. This new file structure also allows users to extract a subsample of individuals or families and the variables of interest to create a substantially smaller file to work with from the beginning of the data analysis process.

Moreover, several special files (called Special Supplemental Files), each with detailed information about a particular topic collected over the years, are released separately, either because the size of the files makes them too cumbersome for storage on the study's main files or because of the unique

nature of the data. Most of these files are public-release files, but some are restricted files that require analysts to sign a special contract with the University of Michigan to assure the confidentiality of the PSID respondents.

Example 4.2 Special Supplemental Files in PSID

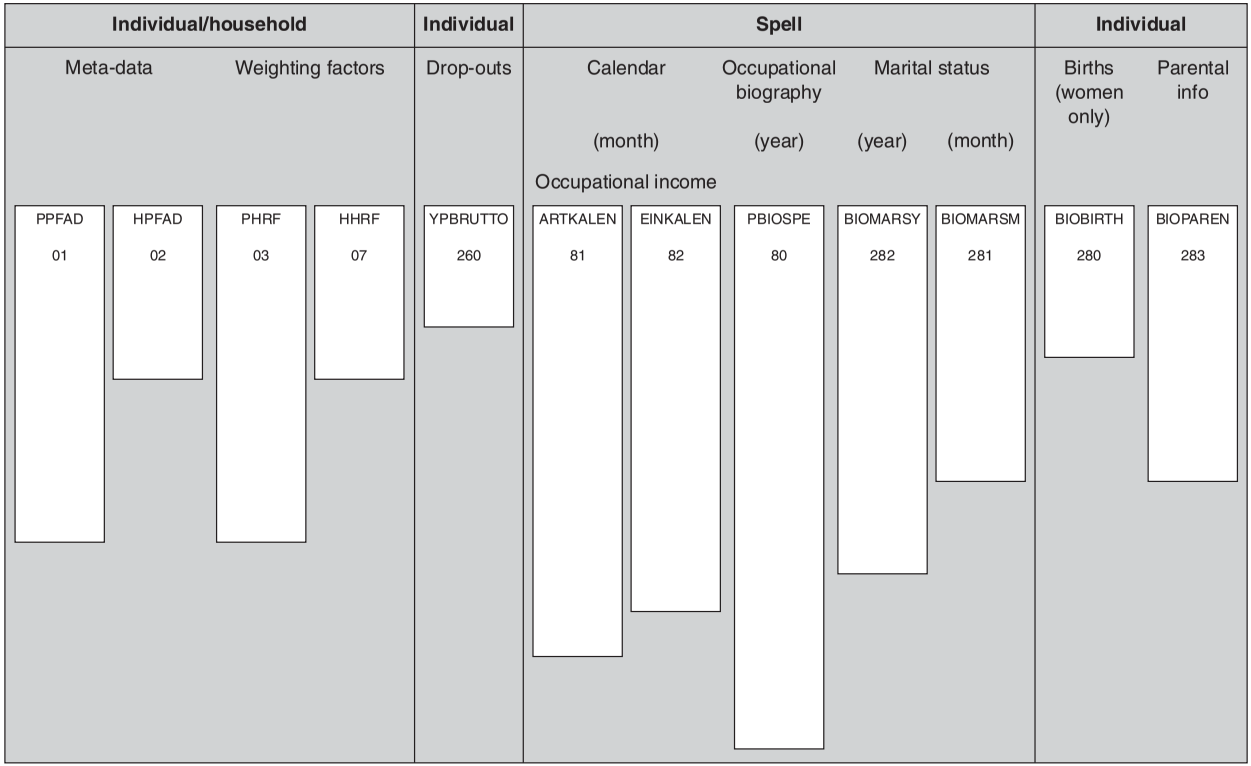
- A newly released Wealth File, which includes data from the 1984, 1989 and 1994 wealth supplements as well as other related information for those years, enables researchers to ask questions about household saving over each five-year period, 1984–89 and 1989–94.
- The 1988 Time and Money Transfers File provides information regarding transfers, in the form of time and money, between a PSID family unit and other persons during the 1987 calendar year.
- A series of health supplements between 1990 and 1995 provide information on health status and health expenditures of the elderly and of their parents. The 1990 Self-Administered Health Supplement contains information about health status, health-care coverage, and long-term care coverage of heads and wives aged 50 and above. The 1990 Telephone Health Supplement contains detailed data on health care costs and utilisation for heads and wives aged 65 and over. It also has information about health services provided or available to the elderly, such as nursing care, transportation, and meals. The 1991 Parent Health Supplement has extensive data about the health status and health care utilisation experience of the parents and parents-in-law of the head of the family. Questions about parents' ability to care for themselves, as well as their housing, income and assets, were included in this supplement.
- The Demographic History Files – the 1985 Ego-Alter File, the 1985–92 Childbirth and Adoption History File and the 1968–85 Marriage History File – provide details about the event and timing of each childbirth, adoption and marriage for PSID family members. The Ego-Alter File also provides retrospective data collected in 1985 on substitute-parenting events and usage of public assistance programmes. Data on these files are structured in a one-record-per-event format to facilitate event-history analysis, and the information is up to date as of the most recent interviewing wave.
- The 1968–85 Relationship File clarifies the crude relationships information available on the main PSID file in early years as well as relating all pairs of individuals associated with a given family. Also included on this file are variables showing co-residence status for pairs of individuals for each year from 1968 to 1985. This file identifies the blood, marital or cohabitation relationships between each pair of individuals who were members of family units that descended from a common, original 1968 family unit.
- The Work History Files (1984–85; 1984–86; 1984–87) contain detailed information about employment and unemployment and the timing of those events.

- The Census Extract Files (1970; 1980) contain a subset of the census data, and the Geocode Match Files contain the identifiers/key variables necessary to link the main PSID data to the census data. This linkage enables data on neighbourhood characteristics within the geographic areas in which panel individuals and families reside to be added to the already rich socio-economic variables collected in the PSID.
- The PSID has gathered substantial amounts of new information about the fact and date of death of many former PSID respondents through efforts to recontact former respondents and through use of the National Death Index of the US Public Health Service (Year of Death File). The resulting information on year of death may eventually be integrated into public-release individual cross-year files.
- As part of its 1990 interviewing wave, and in conjunction with an NIA-funded programme project directed by Lee Lillard, the RAND Corporation, and Linda Waite (now at the University of Chicago), PSID staff asked individuals aged 55 or older who were living in PSID households and who indicated they were Medicare beneficiaries to sign permission forms for access to Medicare claim records between 1984 and 1990. Those who agreed were asked to renew that permission verbally in 1991 until 1995 for Medicare claims made in those years. When combined with questionnaire information on out-of-pocket medical expenditures and the long time-series of core PSID information, the resulting Medicare Record Data should be quite valuable for a number of studies on the health and well-being of the elderly.

Web site: http:/www.isr.umich.edu/src/psid/overview.html

Because of the longitudinal nature of panels, the data are constantly being updated and changed (Freed Taylor, 2000). Some of this is the result of retrospective editing, e.g. where the data collected at a later wave replace data which were imputed/attributed in the earlier releases: as is often the case for income data (see 'Timing and error reduction' at p. 86 below for details). Other changes occur when inconsistencies in the data are discovered and resolved after release. In both cases, changes are introduced into the structure of the data that may well cause real problems for users. Users need to understand what inconsistencies have been found and what data have been changed, revised or adjusted. Again, the documentation must contain detailed information on updating and on changes made, to allow the users of previous waves to repeat, or to continue, their analysis with each new release of the longitudinal file.

Lastly, the complexity and the diversity of the structure of longitudinal data, from the point of view of files and of variables, pose considerable problems to those who try to compare household panel survey data. In a work on the impact of gender on the dynamics of poverty which was conducted using GSOEP and BHPS data (Ruspini, 2000b), the complexity


Individual/household
Individual
Spell
Individual
Meta-data
Weighting factors
Drop-outs
Calendar
Occupational biography
Marital status
Births (women only)
Parental info
(month)
(year)
(year)
(month)
Occupational income
PPFAD
01
HPFAD
02
PHRF
03
HHRF
07
YPBRUTTO
260
ARTKALEN
81
EINKALEN
82
PBIOSPE
80
BIOMARSY
282
BIOMARSM
281
BIOBIRTH
280
BIOPAREN
283

Legend

PPFAD contains records for all respondents, nonrespondents and children under 16 years of age.

HPFAD contains records for all households that ever had contact with the survey. These two files include a person identification number and household identification numbers for each year through the most current survey year.

The file PHRF contains individual-level weights for weighting cross-sectional and longitudinal samples.

Household-level weights are included in the file HHRF.

Information on individuals who drop out of the survey due to death, move abroad or move to another household is located in the file YPBRUTTO.

ARTKALEN, EINKALEN and PBIOSPE are 'spell files' which looks at spells of activity, such as work, retirement and schooling, over each GSOEP survey year. In these files, the unit of observation is not the individual or household, but rather a particular spell or event.

BIOMARSM and BIOMARSY also contain spell information on marital status of each respondent. BIOMARSM is structured on a monthly basis covering the period of time the respondent has been observed in the GSOEP. BIOMARSY is on a yearly basis for the respondent's entire lifespan.

BIOBIRTH (documentation of births per woman) and BIOPAREN (information on parents) are two files with biographical information. There are about 15,000 births documented in the biography file BIOBIRTH, with roughly 11,000 of these persons identified within the GSOEP population. Variables included are: number of children ever born, year of birth and sex of all children not older than 15; personal identifier of children. Variables included in the file BIOPAREN are: year of birth and year of death for both parents; schooling and occupational education of both parents; occupational status of father, when respondent was 15 years old; religious membership of both parents; personal identifier of parents, if they could be identified in the GSOEP.

Figure 4.2 The GSOEP data structure: longitudinal data

Source: Frick, 1998.

of the files and diversity of the structure of the data created marked problems for comparative analysis. More specifically, it was difficult to use the GSOEP data as it necessitated a complex task of re-coding the data. This was because major changes had taken place in the structure of the variables, which had often changed both values and name from year to year. Indeed, while in BHPS data, file variables are named independently of the position they hold in the questionnaire and remain the same over the years except for the first letter of the name of the file (a for 1991, b for 1992, etc.), in the GSOEP file, the variable depends on the number given to that question in the questionnaire (that is on their position within the questionnaire) and because the structure of the questionnaire changes from year to year the name of the variable changes too (see Table 4.3).

Timing and error reduction

Many different considerations must be borne in mind if one is to ensure that the information gathered in a household panel survey will produce high quality data (Duncan, 1989, 1992). Here the concept of 'data quality' is used to mean how well these data are going to be able to satisfy the requirements of those who will be using them for accurate knowledge and information. The accuracy of any data – that is the way in which each datum corresponds to the effective state of the subject/object to which it refers – is the most important attribute of any statistical information (Marradi, 1990; Zajczyk, 1996).

The concept, or the attribute, of 'quality' becomes particularly complex when dealing with longitudinal data. Here 'quality' has many dimensions which include: the nature and the degree of quality of the initial sample; success in following households during the course of the panel; questionnaire design; data processing; data cleaning and minimisation of cross-wave measurement errors (Duncan, 2000). There are, however, some useful and important precautionary measures that may be taken to avoid the risks all panel studies are exposed to, e.g.:[11]

1 the reference population should be clearly identified, with specific attention paid to longitudinal aspects;
2 precise operating rules should be established in order to successfully follow the members of the sample over time;
3 both the questionnaire design and the interview method adopted should be suitable, hence, efficacious;
4 the panel should continue for a sufficiently long period.

In the case of household panel studies, the reference population will change over time because of births, deaths and migrations, thus the panel should be organised, or designed, in such a way that it can adapt to accommodate these events as well as other changes: divorce or separation,

Table 4.3 Principles for naming survey variables (up to eight digits) in GSOEP

Digit	*Meaning*
1	*Wave* (A=wave 1, B=wave 2, … according to West-Samples)
2	Differentiation according to unit of analysis: *P*=individual, *H*=household
3–4	*Number of question* in survey instrument (questionnaire)
5–6	*Number of item* in original question
5 or 7	Differentiation according to sample: *A*=Foreigners, *O*=East Germans
2 to 8	*Text* for Variables in _BRUTTO files and some occupation-related variables in Individual files
5	Differentiation according to green (*G*) and blue (*B*) questionnaire version for old and new households and persons, respectively
Examples:	
AP04	Wave 1; Individual; Question 4
BH0502	Wave 2; Household; Question 5; Item 2
DP24G09	Wave 4; Individual; Question 24; Green version; Item 9
BISCOH	Wave 2; International Standard Classification of Occupation
AP64A	Wave 1; Individual; Question 64; Sample B (Foreigners) only

Source: Frick, 1998 (GSOEP documentation, distributed to users on CD-ROM).

children leaving home and children or parents reuniting, etc. Consequently, the initial sample drawn to make up the panel must be very high quality, particularly because the sample will, in any case, change over time.

As already noted, one of the major problems posed by longitudinal perspective studies is, so-called, attrition. It is important to distinguish between *temporary* and *permanent* attrition: the former refers to a situation when a household, or an individual, re-enters the panel after a period of absence of no more than two waves. Analyses carried out (Winkels and Withers, 2000) indicate that non-response for more than two waves is strongly associated with permanent attrition. Table 4.4 indicates the magnitude of temporary and permanent attrition in the first 17 waves of the Dutch Socio-Economic Panel (SEP).

There is a strong association between non-response and household composition. Usually, people or families with economic problems are the most difficult to contact which poses even greater problems when trying to keep them within the panel over time. Experience with the BHPS has shown that some segments are more likely to drop out between one wave of the survey and the next: families with a high number of members who are unemployed; families who rent their houses; the elderly; widows/widowers and singles; people with a low level of education. In the Belgian Socio-Economic Panel (SEP), the following categories also appeared to have particularly low response rates between the second and the third wave:

Table 4.4 Temporary and permanent attrition in Dutch SEP

Wave	*Number of persons interviewed*	*Temporary attrition (%)*	*Permanent attrition (%)*	*Total attrition (%)*
Wave 1 (Apr 1984)	11,809	–	–	–
Wave 2 (Oct 1984)	11,366	1.6	8.6	10.2*
Wave 3 (Apr 1985)	9,772	1.6	8.6	10.2*
Wave 4 (Oct 1985)	11,838	2.6	6.4	9.0
Wave 5 (Apr 1986)	13,494	3.1	6.6	9.7
Wave 6 (Oct 1986)	14,042	2.2	5.4	7.6
Wave 7 (Apr 1987)	13,577	1.8	5.0	6.8
Wave 8 (Oct 1987)	13,875	1.9	4.0	5.9
Wave 9 (Apr 1988)	13,498	2.2	3.8	6.0
Wave 10 (Oct 1988)	13,772	1.3	3.6	4.9
Wave 11 (Apr 1989)	13,526	1.0	3.6	4.6
Wave 12 (Oct 1989)	13,716	0.4	3.7	4.1
Wave 13 (Apr 1990)	13,404	0.2	4.8	5.1
Wave 14 (Apr 1991)	12,278	0.6	13.1	13.7
Wave 15 (Apr 1992)	13,426	–	–	8.9
Wave 16 (Apr 1993)	13,083	–	–	9.8
Wave 17 (Apr 1994)	13,078	–	–	7.1

Source: Winkels and Withers, 2000: 83.

Note
* Mean figures.

- households from the Walloon part of the country;
- households with two or more people unemployed;
- households with a head aged 75 and over;
- households at the lower (standardised) income levels;
- households which had moved house between the second and the third wave;
- households consisting of young and/or single people (Deleeck *et al.*, 1992).

Moreover, there is almost a pattern to non-responses for certain types of household transitions. For example, changes associated with the dissolution of the original sample households frequently lead to non-response on the part of the members who leave the original home: e.g. when they move out of the parental home or leave a married partner and children (Winkels and Withers, 2000: 95).

The best way to counter the problem of attrition during the period of observation planned, is to ensure that it starts with a high quality initial sample (Duncan, 2000). As has been said, the way the study is designed, the plan for contacting subjects and the guidelines established for how to

follow up the members of the sample over time are all very important. The fundamental rule to set when defining a reference population or populations, longitudinally, is to follow up all the original members of the sample and all those born to these original members. In this way other events, such as births, divorces, children leaving home, new marriages and cohabitations, all of which add new individuals new households and families into the original population, will be reflected in the sample in the same proportion as they are to be found in the general population – unless of course some chance variable comes into play (Duncan, 1992; Ghellini and Trivellato, 1996).

Let us look once more at the experience gained from both the British and the German panels. In the BHPS, the rules adopted to follow up the sample over time and to update the original sample are very clear because they are based on identifying the so-called Original Sample Members (OSM). These OSMs are all the individual members of the families interviewed in the first wave (both adults and any children under 16 who will become members of the panel on their sixteenth birthday);

New members are added to the sample under the following circumstances:

1 the birth of a child to an OSM. Children born to OSMs after the start of the study automatically count as OSMs;
2 the entry of an original member into a new household made up of one or more people;
3 the entry of one or more people into the household of an OSM (Freed Taylor *et al.*, 1995).

Entrants to the sample (categories 2 and 3) become eligible for interview under the standard Office for Population Censuses and Surveys (OPCS) household definition, (i.e. as long as they were living with an OSM and 'either shared living accommodation or share one meal a day and had the address as their only or main residence'). The main requirement for marginal cases of household membership was six months continuous residence during the year. This excluded students who might have been at a parental home during vacations (students were treated as members of their term-time household). The sample for each wave thus consisted of all OSMs plus their natural descendants plus any other adult members of their households, known as Temporary Sample Members (TSMs). It is important to remember that, subsequent to wave one, OSMs were followed into institutions (unless in prison or in circumstances where the respondent was not available for interview, e.g. too frail, mentally impaired, etc.) or into Scotland north of the Caledonian Canal (Freed Taylor *et al.*, 1995).

The rules adopted for following up the members of the German panel are as follows:

- all the people (over 16 years of age) interviewed during the course of the first wave were to be re-interviewed the next year (even when they had moved from one area of Germany to another);
- when children reach 16 years of age they are automatically included in the sample;
- people who enter and become part of a GSOEP household become part of the sample. Starting from the fifth wave (1989), the rule has become to follow up each individual interviewed during the previous wave to collect information about mobility between regions;
- subjects are considered to have left the panel either if they (all members of the household) have not been able to be contacted for two successive waves or if they have refused to take part any more. If a household misses only one wave, then they have to answer a brief, supplementary, questionnaire which serves to gather the information about the previous year which would otherwise be missing.

Furthermore, considerable efforts have been made by the GSOEP to avoid the risk of any under-sampling of problem categories by including a sample of foreign subjects alongside the sample of residents. Notwithstanding, persons with no fixed abode, estimated in 1984 as being about 100,000 individuals, are not represented in the panel; likewise, persons who are in institutions, and who were already underrepresented in the first wave, continued to be poorly represented in succeeding waves. Analysis (Headey *et al.*, 1990; Burkhauser and Wagner, 1990) of the response rate during the second wave (1985) has also highlighted the fact that the rate at which poor people leave the panel is higher than the average rate of attrition. However, in later waves, this difference was less marked.

The measures adopted to successfully follow sample members over time, *tracking techniques*, become increasingly important the wider the time gap is between one wave and the next. Experience gained from what is now a large number of panel studies (Freedman *et al.*, 1980; Burgess, 1989) would suggest adopting the following strategies to maintain a high level of participation/response:

- operations in the field should be planned to favour the participation of potential panel members (e.g., by giving interviewees the chance to decide how their interview should be conducted – face-to-face, telephone or postal, and/or when it will be carried out);
- there should be continuity when placing, pairing, interviewer and interviewee, especially in the early years of the panel. The same interviewer should be sent to the same interviewee: maintaining the same interviewer over waves increases the likelihood of establishing trust, as does the availability of a free phone number for respondents to contact the survey researchers (Singh, 1997);

- having access to the fieldwork agency and continuity of contact is important for respondents, in particular continuity of interviewers across waves. Procedures should be developed to maintain and reinforce the relationship with panel members between waves, e.g. through postal contacts (sending short reports which document and describe how the survey is progressing and explaining how the information gathered is being used; letters that warn of an impending wave; birthday or anniversary cards, etc.);
- monetary or other incentives (e.g. coffee mugs, calculators, lunch bags) should be offered to keep the rate of participation high and to maintain co-operation throughout the duration of the panel. This has been done in both the British and the German panel studies: in the case of the former, the BHPS, a short letter of thanks is sent to each of the interviewees along with a small cheque; while the latter, the GSOEP, enters all interviewees in one national monthly lottery and sends a small gift. Panel survey researchers often send respondents a survey newsletter at regular intervals giving some highlights from the survey findings to generate goodwill for the survey and to maintain contact with respondents (Rose, 2000);
- personalised letters can be sent in order to convince the more reluctant subjects to continue to participate;
- *tracing techniques*, procedures for tracing panel members who cannot be contacted should be developed: among such techniques are asking the Post Office to communicate any changes of address or phone number of the subjects, as well as keeping a note of the phone numbers and addresses of any friends and relations who would be able to offer information as to the whereabouts of the interviewees (Duncan, 1992; Trivellato, 1999).

Example 4.3 Tracking and tracing techniques

The tracking and tracing techniques adopted by the Living in Ireland Panel Survey (LII) were as follows. A personalised letter was sent to each respondent selected in advance of the initial approach by the interviewer. Each household was also sent a brochure which contained information on the survey, discussing in some detail its content and issues of confidentiality, etc. A lottery 'scratch card' was given to each individual who completed the individual questionnaire. Interviewers were instructed to make a minimum of four call-backs to each household in an attempt to make initial contact with its members before the household was dropped from the sample and classified as unavailable (Callan *et al.*, 1996).

The BHPS uses a variety of techniques to keep track of panel members (Laurie, Smith and Scott, 1997):

- providing a named contact person, freephone number and answerphone for respondents;
- recording details of contacts with respondents between interview points;
- passing any relevant information about respondents to the interviewer before each round of interviewing (e.g. news of a family illness);
- an annual pre-fieldwork mailing of a short Respondent Report of research findings and activities with a confirmation of address card for freepost return;
- the inclusion of a change of address card with gift vouchers and thank-you letter post-interview;
- sending a £5 gift voucher as an incentive to any person who returns a change of address card between interview points;
- updating address details between interview points;
- maintenance of an historical record of all addresses ever occupied for each sample member;
- ongoing tracing of respondents both during and between fieldwork periods.

Finally, in the GSOEP, for each successful interview, any respondent (Frick, 1998):

- receives a gift related to the yearly topical module;
- takes part in a monthly nationwide lottery.

Addresses are kept up to date by the fieldwork agency throughout the entire year to monitor residential mobility; e.g. by sending respondents a brochure containing some results of analyses based on previous GSOEP data. The interview situation (face-to-face) ensures a personal relationship, which makes it harder to withdraw from the survey. Maintaining consistency of the interviewer over time is crucial.

To reduce the number of errors that may occur while the information is being gathered the following points should be borne in mind:

- the design of the questionnaire is crucial and close attention should be paid both to the way in which retrospective questions are formulated, e.g. questions based on events are generally preferable to those which are based on calendar dates or periods (see 'Retrospective design and its drawbacks' at p. 96 below). The period of reference should be clearly defined (Dippo, 1989) and care must be taken with the overall organisation of the questionnaire (a way of checking for coherence should be available for interviewers, etc.);
- interviewers themselves should be carefully selected, trained and monitored;
- rules should be established that encourage those directly concerned to answer in person and to establish clear guidelines regarding the use of

surrogates (e.g. only the mother may answer questions which concern the whole family);[12]

- dependent interview techniques (*dependent interviewing*[13]) can be used for variables which are hard to classify, are reasonably stable over time and which should already be both determined and classified during the first contact;
- specific techniques such as Computer Assisted Telephone/Personal Interviewing (CATI/CAPI) can be adopted to administer the questionnaire, techniques which make it possible for interviewers to have access to information gathered in previous waves to improve the longitudinal coherence of the data (Edwards *et al.*, 1993; Trivellato, 1999).

Example 4.4 The CentERpanel

One example of fruitful synergy between innovative data collection techniques and longitudinal research is that offered by the CentERpanel. The CentERpanel is an Internet-based telepanel, representative of the Dutch population, and is made up of some 2,000 households in The Netherlands. Every week, the panel members fill in a questionnaire on the Internet, while at home. Each year about 50 questionnaires of up to 30 minutes each are answered by the respondents in this way. The advantage of such a survey method is that computer-assisted interviewing is combined with panel research: quick results, consistency checks (including time aspects), reliable ways to measure changes and relatively low attrition. Moreover, the results of CentERpanel surveys can be delivered to the client one week after the survey. The CentERpanel was established in 1991, and since then has proved to be of great value for many research projects. Large, complex research projects (like the CentER Savings Project and the Life-cycle project), profit from the possibility of large-scale data collection. Small projects profit from the fact that these telepanel surveys are quick and efficient.
Web site: http://cdata4.kub.nl/website.php?p=meer&l=1#howdoesitwork

---

Other important questions that affect longitudinal panel design are: the length of the panel, i.e. the whole period of observation; the length of the period of reference, i.e. of the retrospective information gathered on each occasion; the number of waves and, lastly, the gap between one wave and the next. When making such decisions three factors should be taken into account, first, the aims of the survey itself, second, the methodological problems – especially how to deal with inaccurate answers – and, third, aspects (and limits) concerning both organisation and costs (Trivellato, 1999).

On the whole, it is reasonable to suppose that the aims of the survey will determine the length of time the panel will run, if only to establish the minimum time it will last, which helps estimate more accurately the dynamics that are of particular interest in the phenomena to be analysed. In general,

the longer the panel lasts, the greater is the wealth of data obtained for longitudinal analysis. For example, the longer the panel, the greater the number of spells of unemployment or spells of poverty completed before the end of the study.[14] The longer the panel, the greater the problems of maintaining a representative sample at later waves because of sample attrition and difficulties in updating the sample for new entrants to the population (Rose, 2000: 41).

Methodological and economic factors predominate when deciding about the number of waves and the length of time period each interview will deal with. The choice will usually lie somewhere between two extremes, on the one hand (almost) continuous interviewing – that is, a plan for frequent, closely-spaced, interviews – and, on the other, a single cross-sectional retrospective interview (or only a few waves carried out at long intervals). The advantages of continuous observation are clear as it helps reduce the number of errors that are caused by faulty memory. Memory errors are nondeliberate errors in reporting of a behaviour, caused either by forgetting that the event occurred or misremembering some details of the event (Sudman and Bradburn, 1982). In the past decade the considerable impact that errors due to poor memory have on studies has been well documented (Kasprzyk *et al.*, 1989; Biemer *et al.*, 1991; Schwarz and Sudman, 1994; Taris, 2000) and the likelihood of such errors occurring increases as the period of reference lengthens. However, the problems associated with a high number of waves have also to be taken into account: the costs; inconvenience for the interviewees – which may lower their propensity to collaborate; *panel conditioning*; the emergence of distortions associated with developing data over time which is emblematically summed up in the so-called 'seam effect'. The seam effect refers to a common phenomenon found with panel data that the levels of reported change between adjacent sub-periods (e.g. from one month to the next) are much greater when the data for the pair of sub-periods are collected in different waves than when they are collected in the same wave (Kalton and Citro, 2000).[15] This is caused by response errors, either misplacing the beginning or end of a spell, or completely forgetting a spell. As Cotton and Giles wrote (1998), it is important to reduce these errors as much as possible so that the measurement of transitions from one state to another is not seriously distorted.

The spacing between the waves is an important decision too, as results tend to change with varying periods of time between the waves of a study (Sandefur and Tuma, 1987). The best combination of waves and of reference periods will depend, obviously, not only on the aims of the study but also on careful weighing up of the pros and cons listed above. In general, at least in the socio-economic area, the quality of the information obtained is closely related to the length of the time gap, the shorter the better, between waves, with periods

of reference (hence distance between the waves) kept at somewhere between a few months and one year (Ghellini and Trivellato, 1996).

Notwithstanding all these precautions it is almost impossible to entirely resolve the problem of non-response. Missing data occur because, for one reason or another, individuals either refuse to participate or they are not willing to answer certain questions. In panel surveys, non-response can be a severe problem because some attrition from the sample occurs at each wave. As already discussed, the cumulative effect is amplified in later waves (Waterton and Lievesley, 1989).

Example 4.5 Types of missing data

Missing data in surveys are generally considered to be of two types: unit non-response and item non-response (Lepkowski, 1989: 348). Wave non-reponse also occurs.

Unit non-response occurs when no data are obtained for a sampled unit because of refusal or non-contact by a sample member.

Item non-response refers to missing data on some items where data should have been supplied by a respondent who has otherwise supplied responses, as when, for example, a respondent may have refused to answer a particular question. It occurs when a sampled member participates in the survey but fails to provide acceptable responses to one or more of the survey items.

Wave non-response (or temporary non-response): this is a form of non-response unique to panel surveys. Wave non-response occurs when data for a panel member are completely missing for at least one wave but present for one or more of the other waves. It thus refers to households or individuals which usually do not respond to one, or sometimes more, waves but subsequently participate in further waves. Circumstances such as an illness may be the cause of temporary non-response. There is a tendency for wave non-response to increase with the age of the panel, as is the case of the non-response which occurs at the initial wave of a panel survey. Often, no attempts are made to contact initial non-respondents at subsequent waves of the panel. Thus, they become total non-respondents for the panel, providing no data for any wave (Kalton and Brick, 2000; Rose, 2000).

There are three basic compensatory strategies that can be adopted to counter the effects of non-response (Lepkowski, 1989: 348–9; Rose, 2000: 18).

Compensating for unit non-response is typically done by weighting. Weights are used to compensate for unequal selection probabilities and for data that are missing because of total non-response and non-coverage. Statistical weights are based on respondent characteristics and adjusted to take account of unit and wave non-response (Kalton and Brick, 2000). For example, in the case of the Dutch SEP (Lemmens, 1991), the sample is

weighted for the following variables: sex, age, marital status, province of residence, degree of urbanisation, region. These are data whose distribution within the population is known.

The second strategy is that of intensive investigation of possible non-response bias by, for example, comparing the responses of continuing respondents with those of non-respondents for questions asked of each group in earlier waves.

Lastly, data must be imputed in the case of item non-response. Imputation means assigning a value, or a set of values, in place of a missing response or a set of missing responses. For example, if non-response relates to questions on income, imputation will be carried out which enables figures for the net income of households and individuals to be determined (Lemmens, 1991). A variety of imputation methods have been developed for assigning values to missing responses in a manner that takes account of the responses given to other items in the survey, including the widely used 'hot deck' and regression based methods (Kalton, 1983; Kalton and Kasprzyk, 1986; Little, 1988; Little and Su, 1989). Hot deck (a statistical matching approach) provides a means of linking a case with missing data with complete cases on the basis of their matching characteristics. Hot deck usually considers only the order of preference of matching complete cases in assigning the missing value (Fay, 1989: 396).

Retrospective design and its drawbacks

Retrospective data are less expensive than prospective longitudinal data as they are usually gathered during one, single wave. However, they do have certain disadvantages connected, in particular, to the distortion due to the inevitable, often unconscious, selectivity of individuals' memories when elaborating their biographies, in remembering. Long-term retrospective data tend to be more unreliable than prospective data: the longer the recall period is, the more unreliable retrospective data tend to be.

- One disadvantage is linked to the quantity of information that an individual is able to remember on one occasion (i.e. when the retrospective interview is carried out). Many subjects simply forget things about events, feelings, or considerations, and even when an event has not been wholly forgotten, they may have trouble recalling it (memory loss and retrieval problems).
- In general, the quality of the data diminishes the further back in time the interviewee is asked to go. There are two main problems with memory: the *omission* effect – when some events that could be important for the study are not revealed, and the *telescope* effect – when the time at which the event took place is not remembered accurately. One particular

type of memory error occurs when respondents omit relevant pieces of information. Respondents may be unable to recall a particular item, or they may be unable to distinguish one item from another in their memory (Linton, 1982). Even if all relevant events have been correctly remembered, if asked *when* they happened, respondents tend to report events as having taken place more recently than they actually did (forward telescoping). The inverse may also occur: some subjects place events further away in the past than they actually happened (backward telescoping). Generally, people tend to assume that distant events happened more recently than they actually did, whereas the reverse applies to recent events (Schwarz, 1996; Schwarz and Sudman, 1994; Taris, 2000). One particular case of the telescope effect occurs when interviewees tend to link the event being studied to certain particular periods (Billari, 1998). Thus only a period that has a well-defined limit, usually the preceding wave, should be used: this helps to reduce the effects of telescoping and, to some extent at least, to keep a check on them (Sudman and Bradburn, 1982; Janson, 1990).

Example 4.6 Memory errors

The factors that influence the ability to remember past events have been studied by various psychologists of memory (Eisenhower *et al.*, 1991; Berrington, 1995). Those that have been identified are (Billari, 1998):

- the time elapsed since the event ('memory decline' effect);
- the importance of the event within the interviewee's life ('event importance' effect);
- the increase in the amount of information sought ('interview difficulty' effect);
- possible interference from memories of other events of a similar nature (inability to distinguish between events; and potential conflicts with information received);
- the social desirability (or emotional content) of the event;
- the psychological state of the interviewee.

Ghellini (1994) carried out an interesting experiment in this field which revealed the consequences of memory inaccuracies in a study on consumption spending. This study analysed the probability trend of memory within six population subgroups that were identified by the combination of age (elderly and not elderly) and classes of equivalent spending (sub-divided into thirds). The memory effect was greater in the elderly, especially among those of higher income groups, who tended to radically over-estimate their past spending. However, quite the opposite, the tendency to underestimate spending, was most marked among those youth who were part of a low-spending group:

> One could indeed suppose that people do not, in reality, remember what they really spent, but rather when estimating it retrospectively, tend to be influenced by their current conditions and consumption (…) thus it could be hypothesised that interviewees used an anchoring technique as an inferential strategy, that is, they were using a current datum, which has been gathered beforehand, as the starting point from which to estimate the retrospective datum, consequently, this latter is closer to their present reality than it is to the reality of their past.
>
> (Ghellini, 1994: 11; Bradburn *et al.*, 1987)

- Furthermore, the way in which individuals interpret their own past behaviour will be influenced by subsequent events in their lives. Subjects tend to interpret and re-interpret events, opinions and feelings so that they fit in with the subjects' own, current perceptions of their lives and past lives, and constitute a sequence of events that 'bears some logic'. This tendency has been called 'modification to fit a coherent scheme' (van der Kamp and Bijleveld, 1998).
- Retrospective questions concerning cognitive and affective states and attitudes are particularly problematic because it is very difficult for interviewees to remember accurately the changes related to particular states of mind, how long these states lasted and the precise order in which they took place.
- In some other areas of interest – such as income, or state of health – it is quite difficult to collect information retrospectively (e.g. information about monthly earnings, blood pressure, weight loss or gain, etc.).
- Like panel studies, retrospective studies too are subject to distortions which are caused by changes within the sample, changes brought about by death, emigration or, even, a refusal to continue with the study.
- Lastly, the length of time required for each interview may also be a major problem. Interviews usually last from one to two hours: the longer the interview the 'richer' will be the information obtained about the life of the interviewee (Billari, 1998).

Sudman and Bradburn (1982), Loftus and Marburger (1983) and Dippo (1989) have suggested a variety of procedures that can be used in order to reduce the distortion that may result from errors of memory. The main ones are:

- aided recall;
- event markers strategy;
- preference for strategies based on episodes/events, rather than those based on calendar dates or on periods;
- bounded recall, i.e. the use of restricted periods of reference.

The first of these four procedures involves providing the interviewee with inputs which will help to improve the quality of what they remember. The 'tools' usually used to aid recall by integrating and enriching retrospective questions are: figures, pictures, lists, copies of newspaper articles and household inventories. Another strategy that can be used to help respondents recall relevant instances of past behaviours is to provide appropriate recall clues. For example, Schwarz (1990) found that when respondents were asked how often they had eaten dinner in a normal or a fast food restaurant, they reported on average 20.5 instances for a three-month period. This rose to 26 times when Schwarz offered recall clues by specifically asking about the number of times respondents had had dinner in different types of restaurants (Chinese, Greek, Italian, Mexican, American).

The *event markers* strategy is particularly useful for limiting and checking on telescope effects. This entails giving a temporal context to the question, by explicitly referring to important events that have occurred in the life of the interviewee: e.g. 'before last Christmas' or 'immediately after your divorce came through' or 'after the birth of your first child'.[16] The idea is to anchor the reference period with salient personal or public events, so-called 'landmark events'. For example, Loftus and Marburger (1983) used the eruption of a volcano to anchor the reference period, when they asked respondents whether they had been victims of crime since the eruption. They also showed that landmarks such as 'Christmas' or 'New Year's Eve' helped to increase accuracy (Taris, 2000: 10). Usually one should be very careful with questions that refer to habitual behaviour. For example, questions such as 'How regularly do you read the newspaper?' run the risk of obtaining an answer that is based on 'how it should be', on the image of themselves that subjects wish to transmit rather than on 'how it is', i.e. on their real behaviour. Thus it is a good idea to restrict this question by placing it within a specific time context, e.g. 'in the past two weeks'. Focusing on a precise period of time makes it easier for the interviewee to remember and makes it harder for 'ideal' behaviour to mask 'real' behaviour.

With strategies based on episodes or events (as alternatives to those based on time periods), if the aim is to measure the length of time a period of unemployment has lasted the researcher can adopt any one of three different strategies. First, the interviewee could be asked 'What was your employment situation on the first day of *x* month/week?'; or a direct question 'How long have you been unemployed?'; or, lastly, a direct question regarding the precise calendar date of the unemployment event: 'When did you lose your last job?'. In the first case the sequence of months or weeks is revealed, retrospectively, backdated from the present; in the second the length of time an ongoing situation has continued is elicited; and, in the third case, there is a date, the date of the event. The results obtained are different. What emerges, e.g.

from employment histories, is that questions based on events are less subject to the clear influences of the telescope effect because they are directly dependent on the salience of the event itself and, consequently, will have less errors than questions which are based on recalling dates, but which seek to elicit the length of time an ongoing episode has lasted. Generally, the question 'How long have you been unemployed?' suffers heavily from the tendency to give rounded figures, or from heaping, which is found in abnormal concentrations for certain periods of time (Torelli and Trivellato, 1993; Trivellato, 1999).

If the strategy chosen is based on the duration of an ongoing episode, then it is a good idea to adopt a period of bounded, i.e. limited, recall – usually the date of the last completed interview is used as a time reference. Events reported during the previous wave can be set aside, which allows the interviewee to concentrate only on what has happened between the penultimate wave and the current wave (Neter and Waksberg, 1964). This helps reduce the telescoping effect as, to some extent, it offers a way of controlling the very human tendency to telescope events when remembering.

Costs and timing of longitudinal research

Longitudinal studies tend to be more costly than trend studies (in terms of time and of the personnel required), mainly because the former are more complex. Consequently, longitudinal studies are usually only carried out by large research organisations (Bailey, 1994) and they need national and often governmental support.

To give some examples:

- The National Child Development Study (NCDS) was initially sponsored by the National Birthday Trust Fund and the Royal College of Obstetricians and Gynaecologists; follow-up studies have been undertaken by the National Children's Bureau and the Social Statistics Research Unit, City University, now known as the Centre for Longitudinal Studies (CLS) and based at the Institute of Education, University of London.
- The National Longitudinal Surveys (NLS) is sponsored and directed by the Bureau of Labor Statistics, US Department of Labor.
- The Survey of Income and Program Participation (SIPP) was originally envisioned as a jointly funded effort by the Census Bureau and the Department of Health and Human Services (HHS).
- The GSOEP is independently funded through the Deutsche Forschungsgemeinschaft or German National Science Foundation (DFG) and located at the German Institute for Economic Research (DIW) in Berlin.
- The first three birth cohorts of the German Life History Study (GLHS) were surveyed within the framework of the Special Research Unit of

the German National Science Foundation that allows projects to be funded for up to 12–15 years. Most of the other surveys within GLHS, and almost all of the data analyses, were organised and financed by the Max Planck Institute of Human Development in Berlin.

Furthermore, while in-depth research tends to use smaller samples – as in the case of the sample used by Elder (1974) – studies that use larger samples tend to be carried out over relatively shorter periods of a person's life-course: this is true in the case of the panel conducted in Lorraine (Panel des Ménages Lorrains), which began in 1985 and ended in 1990, after the sixth wave had been completed (see Appendix 2 for details). Another example, from the study of poverty, is that even though exploration of the mechanisms by which poverty and deprivation are transmitted is of great interest, the prospective longitudinal files currently available are still not sufficient for accurate analyses of the life-courses of both parents and children to be carried out, mainly because not enough waves have yet been completed to permit inter-generational analyses of deprivation.

The higher costs of longitudinal studies are derived from the fact that researchers have to *follow* the subjects over time: have to track people who move house, who form a new family, who move to another municipality. Co-operation has to be ensured in all subsequent waves, as subjects who refuse or are lost by others who did not participate in the previous occasion cannot be replaced. Apart from the actual research costs themselves (*tracking and tracing techniques*), the organisational costs of longitudinal research are tremendous: not only must it be ensured that the same subjects can be traced repeatedly over their life-course, but the research team must be kept constant over the duration of the study (van der Kamp and Bijleveld, 1998).

In the case of the BHPS, a single wave costs more than one million pounds sterling (without counting the costs of the infrastructure). The ESRC Research Centre on Micro-Social Change (now the Institute for Social Research) of the University of Essex was specifically set up to conduct panel studies.[17] Preparations for a BHPS wave, which require about one-and-a-half years and involve six people working half-time, cost about £450,000. The annual cost of the GSOEP runs at about DM4,000,000, of which DM900,000 are spent on the research group involved.[18] Lastly, the approximate cost for the 1998 SIPP panel was US$30,174,000.

The estimated cost of each interview have been calculated at about €150 for the GSOEP, €155 for the BHPS while the Europanel costs only about €65 per interview.[19] The reasons behind these differences are: the ECHP questionnaire is relatively shorter than the other two; both the BHPS and the GSOEP use external research organisations to carry out interviews in the field, while the ECHP is based on information provided by national statistics offices; and, lastly, differences in the way the operations are organised

in the diverse countries as well as differences in the cost of labour in the countries involved (Ghellini and Trivellato, 1996).

To reduce field costs in longitudinal research, many sponsor agencies have approved designs which permit data collection by telephone. As we have seen, the GLHS adopted the CATI technique for interviews; the Dutch CentERpanel is an internet-based telepanel, where every week the panel members fill in a questionnaire on the Internet, at home. Moreover, in the PSID information is collected by means of telephone interviewing and in the HUS panel all 1998 interviews were done by telephone. Finally, all interviews in the Swiss Household Panel (SHP) are made by means of the CATI method (see Appendix 2 for details). Empirical evidence suggests that such changes in mode may not produce biases in the statistics obtained: Benus (1975) noted that data collected by telephone and by personal visit for the PSID were quite similar. Groves and Kahn (1979) found overall that univariate distributions and bivariate relationships were not significantly different for 200 questions administered by telephone and in person. Furthermore, even if telephone interviews have often been regarded as inappropriate for demanding interviews, particularly those with sensitive questions, there is a wealth of actual experience in dealing with long telephone conversations and psychologically sensitive topics (see Brückner and Mayer, 1997 for details). Finally, the CATI system allows researchers to have a continuous, real-time record of interview results, thus enabling them to communicate with the interviewers while the fieldwork was still being conducted (Brückner, 1995).

However, evidence also shows that telephone interviews elicited more rounded financial figures and less detailed responses to open-ended questions (Groves and Kahn, 1979). Moreover, telephone respondents tend to give more 'don't know' answers. This may be related to a difference in perception of length: respondents tend to perceive telephone interviews as longer than personal interviews of the same length and, thus, may be more eager to bring the interview to a close. Consequently, minimising respondent burden seems particularly crucial for interviews conducted by telephone (Federal Committee on Statistical Methodology, 1985).

Notes

1 For further study see Davies (1994).
2 See also Magnusson and Bergman (1990).
3 For a comparative analysis of attrition in HPSs see Singh (1995).
4 See Bailar (1989); Silberstein and Jacobs (1989); Corder and Horvitz (1989); Waterton and Lievesley (1989).
5 For most phenomena reported in surveys, panel participation mainly affects the way in which behaviour is reported – that is to say, responses – while it does not affect behaviour itself (Trivellato, 1999).

6 Campbell and Stanley (1963) identified two criteria for evaluation: internal validity and external validity. An assertion is internally valid if it is based on empirical evidence found within a sample. In univariate analysis, for example, an assertion of the type 'among young people the rate of alcoholism is 20 per cent' only has internal validity if it is confirmed by analysis of the relevant data. In bivariate analysis there is internal validity only if, at the causal level, it is possible to link a specific variation of the dependent variable to variations in the independent variable. In this type of experimental situation, there is internal validity when all disturbance factors that could influence the causal relation are kept under control. This definition can be extended to any assertion, even outside the experimental context (e.g. the internal validity of assertions such as 'there is a positive correlation between the rate of unemployment and the rate of alcoholism'). Hence, internal validity is a criterion that evaluates the extent to which one can believe any given assertion. A piece of research has external validity if its results are not only valid under the specific circumstances in which it was carried out, but are also valid in other situations, i.e. it can be generalised. The external validity of an assertion, just like its internal validity, depends on the techniques used to gather the data upon which the assertion has been based.

7 As already mentioned, longitudinal data already contain information about: the effect of the age of the individual; the cohort effect, linked to the moment at which the individual was born; the period effect, which depends on the moment in time in which the data were gathered.

8 Notwithstanding, the two authors maintain that these data constitute a better approach than would a cross-sectional approach since they make it possible to see the processes that govern life-courses.

9 Those rules which are designed to follow-up and to update the initial sample, so as to ensure that on every wave the sample remains cross-sectionally representative of the population (Trivellato, 1999).

10 Web site: http://www.diw.de/english/sop/index.html.

11 See Kasprzyk *et al.* (1989); Duncan (1992); Kalton (1993); Ghellini and Trivellato (1996); Trivellato (1999).

12 In principle there are no interviews with surrogates in the GSOEP: if the person of reference (woman/man) is not available, then the/a third member of the family is interviewed. The person of reference (who is also asked to fill in the family questionnaire) is the person who knows the conditions of the family nucleus best.

13 With panel surveys, there is the possibility of feeding their responses at earlier waves of data collection back to respondents. This procedure can secure more consistent responses across waves (Kalton and Citro, 2000). *Dependent interviewing* denotes an interviewing strategy: (1) that on the first occasion elicits information about the phenomenon of interest (e.g. profession) through a series of articulated questions that make it possible to classify the subject accurately; (2) in subsequent waves it is then only necessary to ask one question, designed to confirm or not, the classification previously made (e.g. 'Are you still doing the same job as you were doing *x* weeks ago?'), and the whole battery of questions regarding work must only be asked again if the interviewee has changed job. The main reason for using this technique is obvious, it reduces the number of questions. This is, however, a low price to pay if one compares the 'costs' of doing an *ex novo* classification each time but in a less detailed way: a less in-depth classification could, in the case of a variable such as profession, produce confusion, because the subject may well describe his/her occupation in different ways at different times thus giving a false impression of mobility: apparent mobility could be wrongly interpreted as being real mobility (Trivellato, 1999). The ease with which dependent interviewing can be applied depends on the length of the interval between waves and the mode of data collection (Rose, 2000).

14 For example, a spell of poverty has been defined (Bane and Ellwood, 1986) as beginning in the first year that income is below the poverty line after having been above it, and as ending when income is above the poverty line after having been below.

15 This occurs when retrospective data referring to a sub-period within the overall period of reference are gathered – e.g. data required on a month-by-month basis but which is collected within a four-month period of reference – a fairly common practice in panel studies. It has frequently been shown that transitions between the months covered by the retrospective interview are much more contained than is the transition that acts as the 'seam', the join between the next waves. This means that the subject tends to 'flatten', to underplay, the dynamics of episodes when going back in time. This aspect is important especially when decisions must be made regarding both the number of waves and the length of the retrospective periods that will be covered (Martini, 1989; Trivellato, 1999).

16 It is clearly important to immediately ask for information that is required in order to ask the next questions. It is often useful to obtain, immediately, any information that will be required for the next question, especially so, when dealing with questions about matters that may be difficult to remember or that concern other members of the family (e.g. the names of children, etc.).

17 One example: there are about 50 people on the staff of the BHPS working in different units:
- The *Directorate*: seven persons.
- The *Research Group*: 23 persons, economists and sociologists. The idea being to represent these two research disciplines equally and to encourage mixed work groups.
- The *Survey Group*, which is almost exclusively concerned with the technical questions involved in the survey. The group is made up of nine persons who work constantly with the *fieldwork agency* that both gathers and records the data. The technical personnel have to co-ordinate the above mentioned operations, organise training of interviewers, carry out quality control checks (on data and on interviews) and, above all, supervise all activities related to keeping in contact with the members of the panel between waves.
- Lastly, there is also an Information Group (12 people): the *computing manager*; four *computing assistants* and the library staff who are specialised in the literature in longitudinal research and who collect and distribute all the publications and documentation relating to the BHPS.

18 Made up of: a director, eight senior researchers and two clerks (secretarial, administrative). Eight to 10 students and junior researchers (pre- and post-doctorate) are also involved in the research activities.

19 This estimate (Ghellini and Trivellato, 1996), which only refers to operations carried out by NSOs, is based on the following: (a) Eurostat contribution per family of the subject sample: €100; (b) the hypothesis that EC funding covers 90 per cent of costs; (c) the hypothesis that there is a 90 per cent response rate per family; (d) the hypothesis that there are 2.2 adult individuals (16 years or over) per family.

Part II

Longitudinal analysis

5 An overview of the major techniques needed to perform longitudinal analysis

As has already been shown, the term longitudinal is used very broadly. There are many ways of collecting dynamic data – different ways are suited for different types of research – and the longitudinal term is merely the lowest common denominator of a whole family of techniques designed to identify and reveal many types of social change: from time-series techniques for repeated cross-section data to logistic and log-linear models; from structural equation models to longitudinal multilevel methods; from regression analysis to event history analysis.

This chapter aims to offer the reader a brief overview of the techniques most widely used to analyse longitudinal data. The diverse techniques will not be dealt with in full detail, rather, the main ideas behind each technique will be described. This overview will often refer to existing texts on methodology and on statistics and, for in-depth information, the reader should refer to the available textbooks which deal extensively with these subjects.[1]

Time series analysis for repeated cross-sectional data

Time series analysis is the analysis of changes in variables over time. Indeed, a time series is a sequence of observations which are ordered in time (or space). The term, *time series analysis*, is used to describe any one of the various statistical procedures used to tell whether a change in time series data (data arranged in a chronological order: e.g. the annual suicide or the birth rate in the UK from 1900 to the present) has been caused by a variable that occurred at the same time or is due to mere coincidence (Vogt, 1998).

In time series analysis the time point is the unit of analysis, and trends or events in time are variables of interest (Ostrom, 1978; Markus, 1979). The principal difference between a time series and panel data is that, in the former, observations are usually taken on a single entity (individual, country, firm,

etc.) at a relatively large number of time points; while in the latter, it is the individual, or the household, that is observed – with observations not necessarily equally spaced in time; thus, in panel analysis, the observations are made on many entities but at relatively few points in time (Markus, 1979: 7).

Information for time series is usually collected by means of quantitative observations made at regular intervals (e.g. through repeated surveys), such as in the unemployment index, fertility index, GNP, expenditure for pensions and other nationally aggregated variables. Common sources of time series data for social scientists are repeated cross-sectional surveys such as the Eurobarometer, the General Household Survey and the Family Expenditure Survey in Great Britain; the Indagine Multiscopo sulle famiglie italiane (Multi-purpose Survey of Italian Families) in Italy (see Chapter 2 for details).

Time series analysis has two main aims:

- identifying the nature of the phenomenon represented by the sequence of observations;
- forecasting (predicting future values of the time series variable). Some significant areas of application for time-series forecasting methods include the social sciences, marketing and macro-economics.

Time series analyses can be one of three types.[2]

- Temporal analysis involves *describing a trend over time*. For example, has the birth/migration rate increased over time? Has mortality/fertility declined with time? Did the number of births/deaths decrease in the period from the beginning of 1990 until the end of 2000?
- *Discontinuity analysis* is a simple extension that goes beyond a description and offers an interpretation of the impact of some event. Has the trend in mortality changed *since* immunisation was introduced? Did fertility decline *after* a family planning programme was launched? In such cases time series data include an indicator of a disturbance in time.
- *Time series regression analysis* involves interpreting a set of several time series *in which the timing of disturbances varies by area*, but the processes under observation are otherwise comparable. For example, an immunisation programme or a family planning programme may be introduced in an area in phases. The question that arises is, do areas where immunisation is introduced earlier show more precipitous declines in mortality than areas where children are immunised later?

A number of different techniques can be used to analyse aggregate data over time. In time series analysis it is assumed that the data consist of a systematic pattern (usually a set of identifiable components) and random noise (irregular component) which usually make the pattern difficult to

identify. Most time series analysis techniques involve some form of filtering out noise in order to make the pattern clearer.

Two main problems need to be resolved when analysing time series data: *how to identify trend components and seasonal dependency (seasonality) in the data, and the problem of autocorrelation*, that is, how to deal with the correlation (relationship) between members of a time series of observations, such as weekly share prices or interest rates, and the same values at a fixed time interval later. Both problems have a number of special statistical considerations. Indeed, most time series patterns can be described in terms of two basic classes of components: trend and seasonality. The former shows a general systematic linear or (most often) non-linear component that changes over time and does not repeat or, at least, does not repeat within the time range captured by the data (e.g. a plateau followed by a period of exponential growth). The latter may, formally, seem to be of a similar nature (e.g. a plateau followed by a period of exponential growth); however, this repeats itself at systematic intervals over time. Those two general classes of time series components may coexist in real-life data. For example, company sales can increase over the years but they still follow consistent seasonal patterns (e.g. as much as 25 per cent of the total of yearly sales of each year are made in December, whereas only 4 per cent are made in August).[3]

Dependence in a time series refers to serial dependence – that is, the correlation of observations of one variable at one point in time with observations of the same variable at earlier time points. For data in series, such as GNP and entertainment expenditures, the value of any given datum is largely determined by the value of the preceding datum in the series. Autocorrelation is the serial correlation of residual error terms from observations of the same variable made at different times, e.g. interest rates, errors which result from the fact that the value of a datum at time t in the series is dependent on the value of that datum in time t–1 (or some higher lag). This autocorrelation must be controlled before inferences may be made about correlation with other variables. Failure to control autocorrelation may give spurious results, i.e. they may lead one to think that entertainment expenditures, for example, strongly affect GNP. Many forms of time series analysis seek to identify the type of dependency which exists, then to create mathematical formulae which emulate the dependence, and only then to proceed with forecasting or policy analysis.[4]

It is important to remember that time-series analysis methods are more problematic in social science research than are other methods. Among the reasons for this are: the limited information contained in a single time series and the difficulties inherent in formulating models and interpreting results for aggregate processes.

Among the software packages currently available are: Signal Analysis and Time Series Processing (SANTIS), a package with modern graphics and functionalities; MATLAB software, whose routines are periodically updated; Standards Time Series and Regression Package (STARPAC), a library of about 150 Fortran subroutines for time series analysis and non-linear regression; CB PREDICTOR, a time series forecasting software, which is an Excel-based tool that uses established forecasting methods to help identify and extrapolate trends in historical data.[5]

Structural equation models

Structural equation models (SEMs) are models made up of more than one structural equation, that is, equations representing the strength and nature of the hypothesised relations among the structure of sets of variables in a theory (Vogt, 1998).

The main purpose of SEMs is to test specific statistical hypotheses with respect to the relations between a number of variables. One of the most attractive features of SEMs is that not only the relations between manifest variables, but also those between manifest and unobserved hypothetical (or *latent*) variables, can be modelled. Indicators are observed variables, sometimes called manifest variables or reference variables, such as the items in a survey instrument. A latent variable is an underlying characteristic that cannot be observed; it is hypothesised to exist so as to explain variables, such as behaviour, that can be observed. Thus, latent variables are the unobserved constructs or factors which are measured by their respective indicators: they include both independent and dependent variables. The identification of latent variables, based on their relation to observed indicator variables, is one of the defining features of SEMs.

The structural equation modelling process centres around two steps:

1 validating the measurement model. This is accomplished primarily through confirmatory factor analysis. The basic idea is that the variances and covariances between the variables (variance-covariance matrix) included in a model can be decomposed into components attributable to the various relations among the variables. By testing the difference between the observed variance-covariance matrix and the variance-covariance matrix we expect to observe if our model holds, we can assess to what extent the model fits the data (Bijleveld *et al.*, 1998: 211).

Example 5.1 Factor analysis and variance

Factor analysis: any of the several methods of analysis that enable the researcher to reduce a large number of variables to a smaller number of variables, or factors,

or latent variables. Factor analysis is carried out by finding patterns among the variations in the values of several variables: a cluster of highly intercorrelated variables are a factor. This method is often used in survey research to see if a long series of questions can be grouped into shorter sets of questions each of which describes an aspect or factor of the phenomena being studied (Vogt, 1998).

Variance: is a measure of dispersion. The larger the variance, the further the individual cases are from the mean of all cases: e.g. how many hours one person watched TV in comparison to the average number of hours for all the data. Specifically, the population variance is the mean of the sum of the squared deviations from the mean score.

Covariance: is a measure of joint variance (co-variance) of two or more variables (Vogt, 1998; Norusis, 1992).

---

2 fitting the structural model. This is accomplished primarily through path analysis with latent variables. Thus, SEM is a family of statistical techniques which incorporates and integrates path analysis and factor analysis.[6]

Example 5.2 Path analysis and multiple regression analysis

Path analysis: a kind of multivariate analysis in which causal relations among several variables are represented by graphs (path diagrams) showing the 'paths' along which the causal influences travel. In path analysis, researchers use data to examine the accuracy of causal models. A big advantage of path analysis is that the researcher can calculate both the direct and indirect effects of independent variables; this cannot be done using ordinary multiple regression analysis.

Multiple regression analysis: the general purpose of multiple regression is to evaluate the effects of more than one independent variable on a dependent variable. Regression analysis attempts to answer the question: 'What values in the dependent variable can we expect given certain values of the independent variables?' For example, a real estate agent might record for each listing the size of the house (in square feet), the number of bedrooms, the average income in the respective neighbourhood according to census data, and a subjective rating of the appeal of the house. Once this information has been compiled for various houses it would be interesting to see whether and how these measures relate to the price for which a house is sold. As another example, we can build a linear model in which the person's education is the dependent variable and variables such as mother's and father's education and number of siblings are the independent variables.

One starts by specifying a model on the basis of theory. Each variable in the model is conceptualised as a latent variable, measured by multiple

indicators. Several indicators are developed for each model, with a view to winding up with at least three per latent variable after confirmatory factor analysis. Factor analysis is used to establish that indicators seem to measure the corresponding latent variables as represented by the factors: thus, a particular linkage between observed and latent factors is specified. Parameters (that express a relation between two variables) can be free (the parameter may assume any value in the estimation process); fixed (set to some predetermined value) or constrained (its value is equated to those of other parameters). Imposing constraints gives more parsimonious models and makes it possible to impose structure on the model, and thus to test specific theories. The possibility of using constraining parameters is a particularly useful tool in longitudinal data analysis: this usefulness comes from the versatility it offers when specifying models, where many types of paths between variables that are related in time can be created (Bijleveld *et al.*, 1998: 232).

The researcher proceeds only when the measurement model has been validated. Two or more alternative models are then compared in terms of 'model fit', which measures the extent to which the covariances predicted by the model correspond to the observed covariances in the data. 'Modification indexes' and other coefficients may be used by the researcher to alter one or more models to improve the fit. In practice, much SEM research combines confirmatory and exploratory goals: a model is tested using SEM procedures, found to be deficient, and an alternative model is then tested based on changes suggested by SEM modification indexes.

Structural equation models have been proposed (Alwin, 1988; Fergusson and Horwood, 1988) for the analysis of longitudinal data, including data from developmental studies. As Bijleveld *et al.* (1998) pointed out, the ability to constrain parameters, and thus to test whether or not certain paths between variables exist, makes SEM a potentially useful class of techniques for the analysis of dynamic data. In longitudinal data analysis, researchers want to take into account, and model, the time dependence between the measurements; if the temporal dependence between variables can be specified, so too can a longitudinal model.

For the case of two-points-in-time longitudinal data, the researcher repeats the structural relationship twice in the same model, with the second set being the indicators and latent variables at time 2 (Kline, 1998: 259–64). The researcher also posits unanalysed correlations linking the indicators in time 1 and time 2, and posits direct effects connecting the time 1 and time 2 latent variables as well. With this specification, the model can be explored like any other. A path model is thus created for time 1, to which is added a path model for time 2, and more, as needed. When the model is specified, the researcher also specifies that a given variable in the time 1 cluster is

correlated with the same variable in the time 2 cluster, and that the residual error term associated with the latent dependent in time 1 is correlated with the residual error of the latent dependent in time 2, and so on (Jaccard and Wan, 1996: 44–53). SEM is useful for repeated measures and longitudinal designs because it can handle both the correlated independents and the correlated residual errors that will exist between the latent variables at time 1 and time 2 (or in additional time periods).

However, the usefulness of this approach in longitudinal research has been queried by some researchers (Rogosa, 1995; Freedman, 1987, 1991). As with any other longitudinal design, a common problem is attrition of the sample over time: there is no statistical 'fix' for this problem but the researcher should speculate explicitly about possible biases in the final sample when compared with the initial sample. Finally, SEM cannot itself draw causal arrows in models or resolve causal ambiguities. The theoretical insights and judgements of the researcher are of utmost importance.

LISREL, AMOS, and EQS are three popular statistical packages used for carrying out SEM. The first two are distributed by SPSS. LISREL popularised SEM in sociology and the social sciences and it is still the package of reference in most articles about structural equation modelling. AMOS (Analysis of Moment Structures) is a more recent package which, because of its user-friendly graphics interface, has become popular as an easier way of specifying structural models (Kline, 1998).

Log-linear analysis and Markov models of categorical longitudinal data

Typically, log-linear models are used to investigate the interrelationships among a set of variables that are categorical when there are no assumed directions of causality. Log-linear models are useful for uncovering the potentially complex relationships among the variables in a multivariate/multiway contingency table using a limited number of parameters.

Example 5.3 Levels of measurement of variables

Categorical or nominal variable: is a variable that distinguishes among subjects by putting them into a limited number of categories, e.g. by categorising people into female and male. The particular number assigned to a category conveys no numerical information: the codes just represent categories.

Ordinal variable: in this case, responses can be arranged in a meaningful order (e.g. in terms of increasing/decreasing excitement). However, the actual distance between the numeric codes is difficult to define. Variables such as job satisfaction and condition of health are ordinal variables.

Interval variable: if we can interpret the actual distances between the ordered categories, the variable is measured on an interval scale. However, in the interval level of measurement there is no meaningful point zero, that is, scores can meaningfully be added and subtracted but not multiplied and divided. The Fahrenheit temperature scale is an interval scale: when it is zero degrees outside, there is still some warmth.

Ratio variable: if we can interpret distances and also speak of a zero value, the variable is a ratio variable. The ratio scale has an absolute/true zero, that is, not an arbitrary point. Height, weight, age and income can all be measured on a ratio scale. Zero income means no income at all (Stevens, 1946, 1951). Variables that can be measured on an interval or a ratio scale can be defined as continuous, since they can be expressed by a large (often infinite) number of measures.

The term log-linear is adopted because these models use equations that are transformed, by taking their natural logarithms, to make them linear. In other words, log-linear analysis – developed to meet the specific needs of sociologists – transforms non-linear models into essentially linear models through the use of logarithms. Logarithms are exponents of a base number indicating the power to which the number must be raised to produce another number: e.g. the log of 100 is 2, because 10^2 (10×10) equals 100.

Log linear models are mainly used in the elaboration of contingency tables (also called cross-tabulations) which contain many variables, some or all of which are nominal or ordinal measurements. Although log-linear models can be used to analyse the relationship between two categorical variables (two-way contingency tables), they are more commonly used to evaluate multiway contigency tables that involve three or more variables.

Cross-tabulation – a way of presenting data about two variables, that is, a table of the dependent variable against the independent variable – is a basic method for analysing data. For example, a researcher may tabulate the scores on a racism index on the basis of categories of respondents' age and gender; one could tabulate the number of high school drop-outs by age, gender, and school district. In these cases, the major results can be summarised in a multivariate frequency table – a cross-tabulation table with two or more variables. As soon as more variables are introduced, however, there are more relationships to be considered. In four- and five-dimensional tables the number of possible relationships multiplies alarmingly (Gilbert, 1993).

Log-linear analysis is a more straightforward way of looking at contingency tables. The basic idea of log-linear analysis is that a linear model is formulated for the logarithms of the frequencies, instead of for the frequencies themselves. Multiway frequency tables reflect the various main effects and

interaction effects that add together in a linear fashion to give the observed table of frequencies.

The basic strategy in log-linear modelling involves fitting models to observed frequencies in the cross-tabulation of categoric variables. The models can then be represented by a set of expected frequencies that may or may not resemble the observed frequencies. With log-linear models, the researcher tries to predict the number of cases in a cell of a cross-tabulation that is based on the value of the individual variables and on their combination. Each of the variables used in the cross-tabulation (e.g. gender, class, health status, etc.) and its interactions, is tested for statistical significance. In other words, the investigator models the frequency in each cell (the natural logarithm of the observed cell frequency) and examines how this depends on the combination of levels of the categorical variables which define each cell, taking into account sample variation. All variables that are used for classifications are independent variables, and the dependent variable is the number of cases in one cell of the cross-tabulation. Thus, log-linear models are similar to multiple regression models.

However, log-linear models (as well as logit, and probit models) extend the principles of generalised linear models to deal better with the case of dichotomous and polytomous dependent variables. They differ from standard regression in that they use maximum likelihood estimation (MLE) to estimate the parameters of the model instead of ordinary least squares estimation (OLS). MLE is preferred in SEM because MLE estimates are computed simultaneously for the model as a whole, whereas OLS estimates are computed separately in relation to each endogenous variable.[7]

Example 5.4 OLS and MLE

OLS is the commonest form of multiple regression: it works by minimising the sum of squared differences between observed and predicted scores of the dependent variable (i.e. by minimising the deviations of the linear estimates from the observed scores). It can be used when independent variables are dichotomous (i.e. coded as having values of 0 or 1), but not when the dependent variable is dichotomous. In multiple linear regression, the interpretation of the regression coefficient is straightforward: it tells you the amount of change in the dependent variable for a one-unit change in the independent variable.

MLE is a statistical method for estimating the population parameters 'most likely' to have resulted in observed sample data. In other words, MLE chooses the value for which the probability of the observed score is the highest, as the estimate of the parameter. The basic procedure in MLE is as follows: for each possible value that a parameter might have, MLE computes the probability that the

particular sample statistic (observed value) would have occurred if it were the true value of the parameter. Then, for the estimate, it picks the parameter for which the probability of the actual observation is greatest. Unlike OLS regression estimates, MLE does not assume uncorrelated error terms and thus may be used for non-recursive as well as recursive models (Vogt, 1998).

While log-linear models were developed to analyse the conditional relationship of two or more categorical values, logistic, logit and probit models serve to extend the log-linear model to allow a mixture of categorical and continuous independent variables with respect to a categorical dependent variable. The log-linear model is very similar to the logistic model. Logistic and log-linear formulations are mathematically equivalent: the logistic model is, in fact, a special case of the log-linear model, and the log-linear model can also be applied to tables with a binary response variable. Thus, the choice between them will usually depend upon the relative ease (or difficulty) of interpreting the results (Dale and Davies, 1994: 41). Logit regression has identical results to logistic regression: both estimate maximum-likelihood logit models and, by and large, they amount to the same thing. Lastly, both logit and probit usually lead to the same conclusions as they are drawn from the same data.

Example 5.5 Logistic regression, logit analysis and probit analysis

Logistic regression is a form of regression which is used when the dependent variable is a dichotomy and the independent variables are continuous, categorical, or both. Logistic regression is popular because it enables the researcher to overcome many of the restrictive assumptions of OLS estimation regression. In logistic regression the parameters of the model are estimated using the MLE method. In logistic regression the investigator directly estimates the probability of an event occurring: logistic regression is used for predicting whether something will happen or not – such as business failure, heart disease – anything that can be expressed as an event/non-event. Thus, the model does not assume a linear relationship between the dependent and independent variables (it fits a special s-shaped curve). The logistic coefficient can be interpreted as the change in the natural logarithm of the odds associated with one-unit change in the independent variable. The odds of an event occurring are defined as the ratio of the probability that it will occur to the probability that it will not (the odds of an event are calculated as the number of events divided by the number of non-events). For example, the odds of getting a head on a single flip of a coin are 0,5/0,5. If the odds of an event are greater than one, the event is more likely to happen than not; if the odds are less than one the chances are that the event will not happen (the odds of an impossible event are zero). Odds ratios are common measures of association for two variables. An odds ratio below 1 indicates a decrease (i.e. a

unit change in the independent variable is associated with a decrease in the odds of the dependent variable). An odds ratio above 1 indicates an increase (i.e. a unit change in the independent variable is associated with an increase in the odds of the dependent variable). Logistic regression produces odds ratios (OR) associated with each predictor (independent) value. The OR for a predictor variable gives the relative amount by which the odds of the outcome increase or decrease when the value of the predictor value is increased by one unit. ORs are commonly used in epidemiological studies to describe the likely harm an exposure might cause. Epidemiological studies generally try to identify factors that cause harm: those with ORs greater than one.

Logit analysis is a type of log-linear analysis similar to multiple regression analysis. It is used for predicting a categorical dependent variable (such as job satisfaction) on the basis of two or more independent variables. In a logit model, the dependent variable is not the actual value of the variable, but the log odds. Logit regression has numerically identical results to logistic regression, but some computer programs offer both.

Probit analysis is a technique used in regression analysis when the dependent variable is a dummy/dichotomous variable. Probit regression is an alternative log-linear approach to handling categorical dependent variables. A typical use of probit is to analyse dose-response data in medical studies. Like logit or logistic regression, the researcher focuses on a transformation of the probability that Y, the dependent, equals 1. Where the logit transformation is the natural log of the odds ratio, the function used in probit is the inverse of the standard normal cumulative distribution function. Where logistic regression is based on the assumption that the categorical dependent reflects an underlying qualitative variable and uses binomial distribution, probit regression assumes the categorical dependent reflects an underlying quantitative variable and it uses the cumulative normal distribution.[8]

Log-linear models can also be used in the causal modelling of data (Goodman, 1973). Just as with all other causal modelling techniques, with log-linear analysis too, the researcher must specify a theoretical model prior to testing the data. In practice, successive models are tested in order to find the 'best' fit. As the variables may simply be perceived as the same variables measured at two time points, log-linear models are applicable in a longitudinal context (von Eye and Niedermeier, 1999). Payne *et al.* (1994) showed that log-linear and logistic models may be used for modelling trends in relationships between categorical variables. One of their particular strengths is that they allow us to check for variations, observed over time, in the distributions of the variables, so that changes in relationships can be assessed net of these variations. This feature has been particularly important in the analysis of trends in social mobility and on intergenerational mobility: in a

case, for example, when the variables in the contingency table are concerned with the socio-economic status of fathers ($t = 1$) and sons ($t = 2$) (Svalastoga, 1959; Bishop *et al.*, 1975; Goldthorpe, 1987). However, in none of these models is time or the time-dependence between measurements explicitly accommodated (Mooijaart, 1998).

There are models that do specify particular dependencies between observed events which occurred at consecutive time points. A recent development in the analysis of longitudinal categorical data is Markov modelling, that is, a class of probability models termed *Markov chains*. Markov methods are used to analyse movements between states, usually categories of an individual-level response variable such as marital status or voting intention. Markov models have been developed specifically for the analysis of longitudinal data and are relevant to the categorical and qualitative variables which are so common in social research (Davies and Dale, 1994: 167).

In Markov models, transitions from one point in time to another point in time are investigated. Markov models may be simple or highly complex (Mooijaart, 1998: 319, 341).

In their simplest forms, Markov models represent a change process that occurs in discrete time and with reference to a discrete state variable, such as vote intention or occupational status (Markus, 1979). In this model it is assumed that there is a single Markov chain where only the most recent occasion is important for predicting the present state. In the single Markov chain model the population is homogeneous, which means that all subjects have the same transition probabilities, that is, all subjects have the same probability of moving from category i at time point 1 to category j at the time point $t + 1$ (for instance, at time point 1 all employed subjects have the same probability of being unemployed at time point 2).

In more complex Markov models it is assumed that there is more than one Markov chain, where each Markov chain corresponds to a homogeneous subpopulation: each subject, or sample unit, belongs to one chain. These models are called *mixed Markov models*. The idea is that each Markov chain may have its own dynamics. This approach applies to the situation in which the sample is subdivided into strata, each with its own distinct set of transition probabilities: e.g. middle class and working class. Even more complex models are possible if the researcher introduces latent variables. In these models we assume that the observed categorical variables are indicators of one or more latent variables and the Markov chain model is defined not for the observed variables, but for the unobserved, latent variables. In the case of one chain these models are named *latent Markov models*; in the case of more chains they are called *latent mixed Markov models*.

Markov models are useful for a variety of analytical tasks. Nevertheless, these models do have some shortcomings which should be taken into account

(Markus, 1979: 20–1). First, although simple Markov chains may provide useful representations of dynamic processes, they do not explain why individuals change over time. They simply describe the probabilities associated with transitions from one state to another. Stratification of the sample on the basis of an independent variable may increase explanatory power, but the procedure is cumbersome when more than one additional variable is introduced. The Markov approach is also restricted by its general inability to deal with measurement error. With the exception of certain models, simple Markov models assume that all observed change is true change; but when the variables of interest are survey responses, observed change will almost certainly contain some unreliability.

Example 5.6 Markov models

Rajulton (1999) distinguishes between Markov, semi-Markov and non-Markov models. The Markov model implies a simple dependency of events: the occurrence of an event of interest depends directly on the occurrence of the preceding event, and only on it. This means that a transition from one state (origin state) to another state (destination state) depends only on the origin state. A Markov process, therefore, ignores the manner in which the preceding event occurred or the manner in which the origin state was reached.

A semi-Markov model is a modified version of the Markovian one. In a semi-Markov model, changes in states depend on the state of origin, as well as on the state of destination (unlike in a Markov model). The occurrence of an event of interest depends on both the preceding and succeeding events, and on the length of duration between the two events. However, the semi-Markov model ignores the number of events that have already occurred: that is, the Markovian condition is still valid.

There is a third, important aspect to be taken into consideration: the order of events, that should be included in the analysis as past history, greatly influences social or individual behaviour. A model that considers the history of events becomes *non-Markovian*. However, Rajulton (1999) shows that it is generally very inconvenient to build models on non-Markovian lines. In practice, when attempts are made to include the past, a non-Markovian scheme is usually reduced to several Markovian or semi-Markovian schemes.

A number of computer programs are available for log-linear and Markov modelling. For example, SPSS feature log-linear analysis: there are two separate commands in the SPSS Advance Statistics Module for carrying out log-linear analysis: HILOGLINEAR and LOGLINEAR (a more flexible but complex algorithm) (Norusis, 1992). The GLIM system (Francis *et al.*, 1993) is particularly attractive because of its flexibility and interactive facilities (see Gilbert, 1993 for further details). The BMDP system (Dixon, 1988) is

also to be recommended – particularly for the analysis of large tables. Another computer program is LCAG (Hagenaars and Luijkx, 1990). Two computer programs through which Markov models for observed and latent variables can be analysed are PANMARK (van de Pol *et al.*, 1989) and LEM (Vermunt, 1993, 1997). LEM is the most general computer program; PANMARK can analyse the more complicated class of the so-called mixed Markov models. Finally, the collection of programs in LIFEHIST, specifically aimed at analysing life histories, includes a program for non-Markov analysis. This program uses the same algorithm as for semi-Markov models but preserves the different sequences of events already experienced in computing the probability of experiencing a succeeding event (Rajulton, 1999).

Multilevel analysis

One useful technique that can be applied to reveal the link between phenomena and micro and macro social processes is *multilevel analysis*, in particular, that of *longitudinal multilevel models* (Plewis, 1994; Hox and Creft, 1994). In general, multilevel models (also called hierarchical linear models) are used for studying structure in hierarchically organised data, where the units of observations at one level are nested in units of observations at a higher level (MacCallum *et al.*, 1996).

Many kinds of data, including observational data collected in the human and biological sciences, have a hierarchical or clustered structure. Populations commonly exhibit a complex structure with many levels, so that patients (at level 1) are assigned to clinics (at level 2); pupils (level 1) attend schools (level 2); while individuals (level 1) may 'learn' their behaviour in the context of households (level 2) and local cultures (level 3). Similar data structures result from multistage sample surveys. Sample designs typically mirror the hierarchical population structure in terms of geography and household membership: to give an example, in a survey of voting intentions, the respondents (level 1) are clustered by constituencies (level 2) (Jones, 1993). Many designed experiments also create data hierarchies, e.g. clinical trials carried out in several randomly chosen centres or groups of individuals. Lastly, longitudinal designs also give rise to multilevel structures, where occasions of measurement are nested within subjects: the variable measured at the lowest level, the occasion of measurement, is *time* of measurement.

The existence of such data hierarchies is neither accidental nor can it be ignored. Individual people differ and this necessary differentiation is mirrored in all kinds of social activity: e.g. when students with similar motivations or aptitudes are grouped in highly selective schools or colleges. In other cases, the groupings may arise for reasons less strongly associated with the characteristics of individuals, such as the allocation of young children to elementary

schools, or the allocation of patients to different clinics. Once groupings are established, even if their establishment is effectively random, they will tend to become differentiated, and this differentiation implies that the group and its members both influence and are influenced by the group membership. To ignore this relationship risks overlooking the importance of group effects, and may also render invalid many of the traditional statistical analysis techniques used for studying data relationships.

A well-known and influential study of primary (elementary) school children carried out in the 1970s (Bennett, 1976) claimed that children exposed to so-called 'formal' styles of teaching reading, exhibited more progress than those who were not. The data were analysed using traditional multiple regression techniques which recognised only the individual children as the units of analysis and ignored their groupings within teachers and into classes. The results were statistically significant. Subsequently, Aitkin *et al.*, (1981) demonstrated that when the analysis accounted properly for the grouping of children into classes, the significant differences disappeared and the 'formally' taught children could not be shown to differ from the others. This re-analysis is the first important example of a multilevel analysis of social science data. In essence, what was occurring here was that the children within any one classroom, because they were taught together, tended to be similar in their performance. As a result, the data provide rather less information than would have been the case if the same number of students had been taught separately by different teachers. In other words, the basic unit for purposes of comparison should have been the teacher not the student. The function of the students can be seen as providing, for each teacher, an estimate of that teacher's effectiveness. Increasing the number of students per teacher would increase the precision of those estimates but not change the number of teachers being compared. Beyond a certain point, simply increasing the numbers of students in this way hardly improves things at all. On the other hand, increasing the number of teachers to be compared, with the same or somewhat smaller number of students per teacher, considerably improves the precision of the comparisons.[9]

Before multilevel modelling was developed as a research tool, although the problems of ignoring hierarchical structures were reasonably well understood, they were difficult to solve because powerful general purpose tools were unavailable. Special purpose software, e.g. that for the analysis of genetic data, had been available for longer, but this was restricted to 'variance components' models and thus not suitable for handling general linear models. Elaborate procedures have now been developed to take such hierarchical structures into account when carrying out statistical analyses. Software developments now allow such models to be applied to a wide range of different structures.

Multilevel analysis began to be developed in the early 1980s, even though the principles underlying such analysis had been laid down 20 years before. It is a particularly useful tool in the field of geographical and educational research. It uses highly innovative techniques which allow researchers to work at diverse levels of analysis contemporaneously. This, in its turn, makes it possible for the *grouping effect* to be taken into account simultaneously. The *grouping effect* is the hierarchical structure that characterises social life – (individuals live in families which are, in their turn, established in areas which are under diverse local authorities; students are grouped into classes, which are part of different schools, which latter are, in their turn, part of different local, county and regional contexts). Multilevel models provide a framework for representing the structure of such data both within and between levels, thus eliminating the need to aggregate data or to analyse different levels separately (MacCallum *et al.*, 1996).

The extent to which this hierarchical structure influences the measurement of the interest phenomenon itself can be tested using multilevel analysis. For example, if we are measuring scholastic progress, the interest variable may be to find out the extent to which this progress (learning) differs between students, between classes and between schools; or, if the study is about poverty or about the efficacy of welfare programmes, then the interest phenomenon could be how experiences vary between families who live in different areas, communities or towns/cities.

Multilevel models are based on regression techniques. Indeed, *multilevel models* could be seen as an extension of conventional regression analysis, which can be applied to data with a hierarchical, *clustered* structure and which allow the inclination of regression lines – that is, the graphic representation of a regression equation – to vary between groups. The key feature of multilevel models is that they specify the potentially different intercepts and slopes of regression lines. These procedures do not fit one single relationship, but allow the relationship to vary from context to context. In this way, both the 'classic' *single-level* approach (in the case of the regression line, one intercept and one slope) and, consequently, the individual/aggregate level dilemma is overcome (Jones and Duncan, 1994). Multilevel models were explicitly developed to resolve this dilemma by working at more than one level simultaneously, thus with the potential to offer improved estimates. In substantive terms, by working concurrently at the individual and contextual levels, these analytical models begin to reflect the complexity of social organisation and are more faithful to the nature of the social world.

This technique is particularly useful when dealing with longitudinal data which have an inherent hierarchical structure. For example, HPSs gather repeated measurements, data, on income and consumption (level 1), relating to diverse individuals (level 2), in different sectors of the economy (level 3)

(Jones, 1993). Indeed, multilevel models can be fruitfully applied to the study of univariate (on a single response variable) change and can easily be extended to the analysis of multivariate change. As MacCallum *et al.* (1997) wrote, multilevel models have the potential to provide valuable information about relationships between patterns of change on different variables. The investigator has a variety of alternative models and strategies. For instance: a) various linear and nonlinear models for representing change on each outcome; b) the separate, as opposed to simultaneous, analysis of multiple outcomes; and c) the potential inclusion of additional variables that might be related to the basic functions of the outcomes.

Even though multilevel modelling is a rapidly developing area of research, it should be remembered that it is generally difficult to learn and carry out. The most widely used package, at present, is MLN. It was developed by the Multilevel Models Project (Institute of Education, University of London), and is suitable for an arbitrary number of levels. A recent development is MLwiN, the latest release of the MLN program: it is a Windows application that provides a visual interface for multilevel modelling. Other packages are: HLM, a three-level software produced by Bryk and Raudenbush (Bryk *et al.*, 1988); VARCL, a three-level software by Longford (Longford, 1988); MIXREG/MIXORR, a software which includes discrete response models.[10]

Event history analysis

The life-course approach has developed a body of techniques in the field of Event History Analysis (EHA). Here, EHA means a set of mathematical models for the analysis of those processes that lead to one single event or to events that may be repeated over time (Mastrovita, 1998). More precisely, EHA is the name given to a wide variety of statistical techniques for the analysis of longitudinal data (event history data) and for studying the movement over time (transitions) of subjects through successive states or conditions, including the length of the time intervals between entry to and exit from specific states (Blossfeld and Rohwer, 1995: 33).

EHA is usually used in situations when the dependent variable is categorical (Carroll, 1983; Tuma and Hannan, 1984; Allison, 1984). However, even changes noted in continuous dependent variables can be dealt with: e.g. the event, 'temperature', could be considered to be an unexpected rise in body temperature (Allison, 1995; Mastrovita, 1998).

As Skinner (2000) wrote: for the simplest kind of event histories, we may suppose that for each individual in the population:

- an initial event occurs at time I_i;
- a terminal event occurs at time $I_i + T_i$;
- there is an association vector of covariates x_i.

Table 5.1 Examples of event histories

Initial event/ origin state	*Terminal event/destination state*	*Covariates*
Start of first spell of unemployment	End of first spell of unemployment	Age, sex, occupation
Birth	First marriage	Sex, social class, education, occupation
First marriage	First birth	Age at first marriage, social class, education, occupation
First marriage	Divorce	Age at first marriage, social class, education, occupation
End of full-time education	First full-time employment	Sex, social class, education

Source: Re-elaborated from Skinner, 2000: 121.

The aim of the analysis may be to study how T_i depends on the value of the vector of the covariates x_i. Some examples of such event histories and associated covariates are shown in Table 5.1.

Example 5.7 EHA models

The most basic event history model is based on a process within only one *single episode* and *two states* (one origin and one destination state): e.g. the duration of a first marriage until its end, for whatever reason. In the single episode case each unit of analysis which entered into the origin state (married for the first time) is represented by one episode. If there is more than one destination state, we refer to these models as *multistate models*. Models with a single origin state but two or more destination states are also called *models with competing events or risks*. For example, a housewife might become 'unemployed' (i.e. enter the state 'looking for work') or start being 'full-time' or 'part-time' employed. If more than one event is possible, that is, if there are repeated events or transitions over the observation period, we use the term *multi-episode models*. For example, an employment career normally consists of a series of job shifts (Blossfeld and Rohwer, 1995: 34).

In EHA the first important concept is the *risk period* (Yamaguchi, 1991). Indeed, it is possible to divide the time period when the event does not occur into two parts: the *risk period* and the *non risk period* that the event will take place (e.g. the birth of the first child).

Another important concept in EHA is that of the group of individuals who risk experiencing the event, the *risk set* within the observation window.

For example, if we are studying a population of 200 individuals, and we are studying the risk that they will change job, all 200 will be considered to be at risk in the first year. If only 11 persons out of the 200 do change their job in the first year, these 11 will no longer be at risk in the second year (they could be at risk for a second job change, but we are only studying the risk of a non-repeated event). Thus the number of subjects at risk drops, each year, by the number of subjects who have undergone the event in that year (Allison, 1984).

The third key concept is that of the *hazard rate* or *hazard function*, that expresses the probability that an event will take place at time *t*, given that that event has not taken place before time *t* and that, consequently, the population could still be considered to be 'at risk'. The hazard function or *h*(t), is defined as the ratio between the probability of the event taking place, *f*(t), divided by the survival probability (or *survivor function*), *S*(t), which is the probability that the event will not take place before time *t*. The *hazard rate*, or *transition rate*, is the fundamental dependent variable (even though it cannot be observed) in an event history model. The risk of an event occurring within a given period is then regressed on a set of explanatory variables, called covariates, some or all of which may themselves vary over time. The term hazard comes from bio-statistics, where the typical event is death. The hazard function may take on very different forms, depending on the type of process that is being studied. For example, the chance that women will experience a first childbirth is zero for the first 12 years of their lives, then increases strongly, only to become zero again around age 45 (Taris, 2000: 101).

There are two main groups of methods that can be used to analyse *hazard rates*: non-parametric, partially parametric (or semi-parametric), or fully parametric estimation methods (Yamaguchi, 1991).

Non-parametric methods (life table method and Kaplan-Meier method) do not specify the relation between hazard rates and explanatory variables/covariates. Instead, separate estimates of rates as a function of time are obtained for distinct strata, such as ethnic groups, which are distinguished by a time-invariant categorical variable. For this reason, they are particularly suited for initial, exploratory, data analyses.

Example 5.8 The Kaplan-Meier method and the life-table method

The most basic methods of event history analysis involve techniques that are analogous to descriptive statistics (Tuma, 1994). The survivor function tells us the proportion of the population that are 'alive' or 'not experiencing the event' at the time. We can often gain some insight into the distribution of survivors over time by plotting the survivor function. We can compare two or more groups' survival rates using this graph approach, along with more formal statistical tests of differences in survival functions. Among the procedures that generate estimates of the survivor function are the Kaplan-Meier method and the life-table. The

Kaplan-Meier method is useful for small datasets or when the time that the event took place was accurately recorded. Furthermore, the procedure tests equality between the survival functions of diverse groups. For example, it can be used to estimate the proportion of employees who will still be working in the same firm at different points in time after they were first taken on. Each time an employee leaves the firm, the proportion of workers who are still there will be estimated: e.g. if a worker leaves the firm after three-and-a-half years, the proportion of workers who are still working will be estimated. The life-table estimates survival functions for fixed points in time, e.g. monthly or yearly: this is the best method to use when dealing with large datasets (because it needs less computing time and space) or when data about time are not accurate. Compared to the Kaplan-Meier estimator, the life-table method has the disadvantage that the researcher must define discrete time intervals: the results, therefore, depend to some extent on these arbitrarily defined time intervals (Blossfeld and Rohwer, 1995; Mastrovita, 1998).

Fully parametric (exponential models, Weibull models, Gompertz models) and partially parametric methods (Cox regression) estimate the effects of covariates on hazard rates. In parametric models the hazard function is assumed to comply with a particular functional distribution. For instance, the hazard may be thought to be the same at all time points (a constant); in other cases the hazard rate is assumed to vary over time. It is thus important to choose the 'right' functional distribution of the hazard rate because the estimates of the effects of the explanatory variables are estimated in relation to the distribution chosen (e.g. the risk of leaving a job would decrease over time). Unfortunately, the researcher seldom knows a proper functional form: choosing the correct distribution is a difficult problem. As in many instances investigators are not particularly interested in the distribution of the hazard function but rather in the effects of the covariates, they often opt for a semi-parametric model (Taris, 2000: 106–7).

The semi-parametric model developed by Cox (1972, 1975) is often more appropriate than a parametric model. The Cox model does not assume any specific distributional shape for the hazard function. This is useful when the investigator does not have explicit ideas about the shape of this function, when the hazard rate is too irregular to fit any particular distribution, or when one is only interested in the magnitude and direction of the effects of the explanatory variables. Thus the semi-parametric approach is, effectively, a generalisation of the fully parametric approach, but it cannot be used when one is interested in the way the hazard rate is affected by time (Mastrovita, 1998).

Another attractive feature of the semi-parametric model is that time varying covariates can easily be included in the analysis. In event history analysis an important distinction is always made between the possible causes

of the events: some of these variables, such as sex, will be constant over time (*time-constant covariates*), while others, such as income, may change over time (*time-dependent* or *time-varying covariates*). These variables create problems for standard statistical procedures such as multiple regression. To give just one example, when analysing the event 'divorce', the covariates that could be used will include characteristics that do not change during the marriage (e.g. race, level of education before marriage, age at moment of marriage) and those that change over time (e.g. income, number of children, work status).

Lastly, unlike regression analysis, event history analysis is able to handle a certain kind of missing data referred to as 'right-censored' data. Since in event history analysis the termination of the entire observation time period is given, an episode may not be closed. By *censoring* we mean a state that occurs when the information about the duration is recorded incompletely because of the temporal limits of the observation window we take into account. Censoring of a time period may occur from the right (observation stops before the event is observed) or from the left (observation does not begin until after the event has occurred, i.e. the correct beginning of a process is unknown). Right censoring affects estimation procedures because the timing of the transition is not observed for one reason or another. One reason for right censoring might be that the event in question never happens to certain individuals. For example, not all people experience first marriages or change jobs. Another reason might be that some individuals have not experienced an event during the period of observation, but may experience the event some time later. In either case, all we know about an individual's event-time is that it exceeds the time they were last observed. Although the data are missing, the individual's censoring-time still constitutes valuable information when estimating transition rates.

The usual, conventional method adopted when analysing censored data is the 'life-table'. This procedure makes the simple assumption that censoring is independent of the attrition process. Whatever observation is gathered from such cases is used in calculating populations at risk up to the time of censoring. The survival process is analysed in small discrete time periods, with simple assumptions made about the temporal distribution of risk within discrete intervals. Survival probabilities are calculated for each interval, so that cases that are censored at some point in time can be used in the denominators of rates for time segments prior to the point of censoring. Discrete probabilities computed in this fashion can then be accumulated multiplicatively to show the implication of a series of probabilities for the overall survival process. The Cox regression model can also be used to analyse data that contain censored observations, whereas multiple linear regression cannot be used for analysis of time-to-event data, since there is no way in which censored observations can be handled.

Thus, despite the practical difficulties and the conceptual complexity, EHA does offer some advantages:

- it allows information associated to duration or timing data to be used efficiently;
- it makes it possible to model time/duration dependence;
- right-censored observations can be dealt with adequately;
- time-varying event predictors can be used;
- it can satisfactorily deal with 'unobserved heterogeneity', this refers to the type of situation where some of the explanatory variables that the hazard rate depends on have not been, or cannot be, observed.

As Mayer and Huinink (1990) stated, event history techniques have revolutionised the analysis of longitudinal data as they make it possible to estimate the impact of a factor or a set of factors on the timing and sequencing of life-course transitions. For example, a researcher might be interested in estimating the effects of individual-level, family-level or community-level traits on the timing and sequencing of marriage, birth, job changing, migration or mortality. Thus, EHA has an explicit longitudinal perspective. The importance of studying events was highlighted by Elster (1989): the ideal aim of the social sciences is to explain the events individuals live through. Explanations of events should be given priority over explanations of states because the state itself is seen as being the result of the events. In Elster's opinion, an ideal explanation is achieved when one or more events can be identified as being the cause(s) of the event being studied (Billari and Rosina, 1999).

Among the currently available statistics software packages are: the GLIM and RATE packages (developed by Tuma in 1979) and BMDP (Dixon, 1988) which is particularly versatile and useful for Cox model estimates. SAS and SPSS are two other statistical packages used for carrying out survival analysis. Lastly, there is a program that has been specifically designed for longitudinal data analysis: TDA (Transition Data Analysis), written by Rohwer. This program is continually being improved and is distributed along with the text by Blossfeld and Rohwer (1995).

Sequence analysis

Event history data consists of sequences of qualitatively different states occupied by the participants during the observation period, as well as the timing of transitions from one state to the other. Such data can be analysed also by examining event histories *as wholes*.

Most of the 'classical' methodological tools used for analysing longitudinal data focus on single events, instead of on mobility patterns, in terms of the serial succession of sequential events. Conversely, *sequence analysis* enables us

to consider and handle the information about whole career sequences, taking into account the information about the *duration and frequency spent in different statuses as well as their location and ordering* (Scherer, 2000). Therefore using this relatively new methodological tool we can treat a career trajectory *holistically* (Halpin and Chan, 1998; Rohwer and Trappe, 1997). This property of sequence analysis has revealed itself to be even more valuable when studying, for example, women's careers, given their rather high rate of unstable and interrupted careers, a feature which cannot be captured if one takes only single snapshots into account (which could mean cross-sectional as well as panel data) instead of dynamic mobility patterns.

Among the areas of sequence analysis research which have been identified as being of particular concern within the social sciences are the following (for a review of the literature see Abbott, 1995):

- job mobility and career processes;
- developmental profiles;
- understanding behaviour;
- decision development in groups;
- the development of social expectations;
- sequencing and social structure in family conflict;
- crime, drug use;
- the evolution of market structure; market leadership.

The basic idea of sequence analysis is to represent each life-course, or trajectory in the life-course, as a 'word', or more precisely, as a string of characters (in some cases numerical). As Billari (1999) wrote, when representing a life-course as a sequence of events, one normally assigns a letter (or a number) to an event, and the ordering of events gives the ordering of letters in the word. If we want to represent union behaviour and one person first forms a cohabiting union (event denoted by *C*), then he/she gets married (*M*), then he/she gets divorced (*D*) and remarries (*M*), a representation of the sequence of events is:

CMDM

The main problem with this representation is that one cannot take into account the distance between events, and it is not clear how to behave when such events happen simultaneously. The approach is, however, interesting when the number of events is low, or the complexity of life-courses is limited. Mainly to overcome the limitations of this approach, and to take the duration between events explicitly into account, research efforts have focused on representing life-courses as (recurrent) sequences of states. The origin of this approach can be traced back to computational biology (Sankoff and Krustal, 1983; Waterman, 1995).

Example 5.9 The properties of sequences

Abbott (1995) discussed a number of the properties of sequences.

First, events in a sequence can be unique or they can repeat. A sequence in which events cannot repeat, is 'non-recurrent'. The length of such a sequence cannot exceed the size of the universe of events (the elements of a sequence are 'events', drawn from a set of all possible events in a set of sequences, the 'universe of events'). A sequence in which events can repeat is a 'recurrent sequence'. The length of a recurrent sequence has no limit, but is typically set by some sampling frame – a lifetime, a wave of data collection.

Second, sequences can show dependence between their states. The most familiar examples of this are stochastic processes, in which the n+1th element of the sequence is some specified function of the nth or perhaps earlier elements. On the contrary, there may be minimal inter-state dependence.

Third, there can be varying degrees of dependence between diverse whole sequences. We sometimes have sequences in which the occurrence of an event in any one sequence prevents that occurrence in any other; e.g. there can be only one President of the European Union, at any one time (White, 1970). This is true in a looser form for phenomena like 'upper-classness' or 'working in the farm sector' where larger constraints, usually conceptualised in sociology as 'constraints on the marginals', limit possibilities across sequences.

Fourth, sequence can be investigated either for itself or as an independent or dependent variable. Sometimes we are interested simply in the patterns in a collection of sequences. Other times we wish to know how a prior event sequence affects the immediate future, like when we try to predict joblessness given a prior sequence of job experiences. Still other times we wish to know what accounts for different sequences of behaviour – what prior variables, for example, lead to poverty?

In this view, one explicitly considers life-course data as being fragmented into discrete time. The assumption is that either there is a 'natural' discrete time unit (e.g. month or year) in the data, or that some 'discretisation' has been performed. As a simple example, we shall consider three states: single (*S*), cohabiting (*C*) and married (*M*), in a monthly scale from 20 years to 25 years. The sequence representation of an individual life-course may thus be (Billari 1999):

SSSSSSCCCCCCCCSSSSSSSSSSSSSSSSSSSSSSCCCSSSSSSSSSSSSSSSSMMMMMM

This person, starting as single on their twentieth birthday, started cohabiting at the age of 20 years and 6 months, broke up the cohabitation at 21 years and 2 months, started a new cohabitation at 22 years and 11 months, broke it up at 23 years and 2 months, and got married at the age of

24 years and 6 months. The representation as sequence of states can be easily reverted into an event history representation with a discrete time scale.

The representation can be further generalised to a set of parallel words. Life-course should be considered as being composed of several parallel domains. This happens for instance when marital history and reproductive history are studied jointly. We may then have a representation where each individual is represented by a vector of states at each point in time. Moreover, this idea can be used when we are interested in the parallel careers of different individuals. Then, the drawback is that the number of states and, consequently, the scale of the alphabet that we need rises quickly (Billari, 1999). One of the big issues is, indeed, how the number of distinct careers can be reduced to a manageable number. As Taris wrote (2000: 122), if careers are observed on eight occasions and participants can belong to one out of four states, the number of distinct careers amounts to 8^4 (as many as 4,096 careers)! Clearly, the number of careers must be reduced. There are different approaches to this problem: very broadly, one approach focuses on the type, number, direction and relative frequency of the transitions that occurred during the event histories of the participants, resulting in a quantification of the careers of interest. Another approach focuses on classification of similar careers: it involves computing distances among careers, using correspondence analysis (see Taris, 2000 for details). It is indeed possible to calculate a measure for the distance between the different individual career sequences, either in comparison to a standard sequence or by comparing each sequence with each other. This distance can be used as an input for a variety of different applications or, simply, for descriptive purposes (Scherer, 1999). *Correspondence analysis* is a descriptive/exploratory technique designed to analyse two-way and multi-way tables containing some measure of correspondence between the rows and columns. The results provide information which is similar in nature to that produced by factor analysis techniques, and they allow one to explore the structure of the categorical variables included in the table.

Example 5.10 Approaches to reduce the number of careers in sequence analysis

Among the approaches proposed to reduce the number of careers, focusing on a social science perspective, there are (see Billari, 1999 and Taris, 2000 for details):

1 *Description based on the features of individual sequences*
Computer graphics – in cases with access to colour representations – are particularly helpful in the description of sequences of states (Rohwer and Trappe, 1997; Rohwer and Pötter, 1999). An interesting example is offered by Scherer

(1999, 2000) in a study of early career sequences, that is, on the way in which men and women enter the labour market in two European countries (Germany and the United Kingdom). In the graphs presented, one line is drawn for each individual based on monthly status information. Colours represent the different statuses (qualified positions, unqualified positions, unemployment, not in the labour market).

2 *Optimal matching analysis*

The 'optimal matching approach' is based on a notion of similarity, or dissimilarity, between pairs of sequences (Abbot and Forrest, 1986): it computes distances between event histories, explicitly taking into account the temporal order of the elements in these careers. This method was originally developed in the biomedical sciences for examining the similarity between DNA and RNA sequences (Doolittle, 1990; Sankoff and Kruskal, 1983; Waterman, 1995). The basic idea of the optimal matching approach is to measure the dissimilarity of two sequences by considering the question of how much effort is required to transform one sequence into the other one. The more alterations are necessary, the greater the difference (and the greater the distance) between these sequences. The distance between two sequences may thus be defined as the minimum cost of transforming one sequence into the other one. As a result, one obtains a distance matrix. This may be employed as an input for every kind of analysis requiring proximity data (e.g. clustering and multidimensional scaling).

3 *Clustering binary sequences*

Billari and Piccarreta (1999) tried to solve the problem of building meaningful groups by using algorithms for clustering binary variables. The algorithm applies to a series of parallel sequences that can be represented by binary variables. For a meaningful interpretation of the algorithm the events must be non-renewable. This is a hierarchical divisive algorithm, which means that it starts from the whole sample that it then divides into two groups. One of the two groups is then split into two subgroups. The procedure can be iterated until each individual belongs to an 'own' group. This is also a monothetic algorithm: each group is divided into two subgroups according to the values of a single variable (binary in this case). To perform the splitting, a single relevant variable must be selected. The splitting variable is selected in such a way that the two subgroups induced by its categories are characterised by the maximum homogeneity within groups and by the maximum heterogeneity between groups. The main advantage of this algorithm is that it provides easily interpretable clusters: the groups obtained are, in fact, perfectly characterised by the presence (or absence) of certain attributes (those measured by the splitting variables). Another interesting feature is that it is possible to identify the most relevant variables in the clustering process (the splitting variables) and to rank these variables according to their importance in the clustering process.

4 *Multiple correspondence analysis of sequences*

Van der Heijden (1987) illustrates the use of multiple correspondence analysis in the study of sequences. This technique is widely used in the analysis of qualitative data within the social sciences. Multiple correspondence analysis is useful in the context of life-course research both in order to synthesise the cross-sectional situation at each point in time, and to analyse the differences between individuals and identify those which are particularly 'distant' from the mean. Graphical inspection is fundamental to this approach. Until now, applications have mostly focused on diaries and time-budgets, that are, however, substantially cross-sectional, and sometimes the time points needed to be aggregated. Nevertheless, this technique can be particularly useful when sequences are generated by non-renewable events and, as we have seen, in such a case, cross-sectional situations themselves depend on the past.

What kind of instruments may be used for data collection when one wants to build up a sequence representation of life-courses? Each instrument that permits event histories to be constructed can be used to produce sequences of states. So, retrospective surveys may – and in fact have been – used to produce sequence data too. It is not surprising that the technique of data collection of life-courses known as the *life history calendar* or LHC (Freedman *et al.*, 1988) was based on the idea of representing life-courses in a fashion similar to the sequence of states. Such methods are considered in the broader spectrum of the collection of 'biographical/life history matrices' (Olagnero and Saraceno, 1993; Settersten and Mayer, 1997) in life-course research (see Chapter 2 for details).

The ideal source for sequence data is a population register – provided, obviously, that it contains the information the researcher is interested in. Such sources have the advantage of providing the same amount of data as retrospective surveys do, without problems of recall and usually with less information missing. However, such sources are rare and costly, and information is usually collected at 'distant' points in time (e.g. every 10 years). Moreover, we can only build up sequences for trajectories that are officially recorded. Record linkage of different census records may provide a sequence that is sufficiently long to be complex enough to need specific techniques: for instance, three censuses linked, each of them asking for the present state (e.g. residential location) and two past states in the inter-census period, will provide sequences of nine time points. Of course, in that case, the information between the measurement occasions is lost.

A further and more widely available source of sequence data is that drawn from panel surveys. Such surveys usually gather information about the states of individuals at several points in time. They do not necessarily provide full event histories, thus discrete-time event history models have often been used for analysis of such surveys.

There is a great deal of software available for sequence analysis in the natural sciences (a review can be found in Abbott, 1997). In the social sciences, there are usually a high number of relatively short sequences, e.g. a sample from the population of interest. Thus, each individual sequence is of little interest, rather, the aim is to gather information about a group of people. This problem, frequently met with in the social sciences, requires specific software but so far only two programs have been developed to study life-courses represented as sequences. One is OPTIMIZE, a program developed by Abbott *et al.* (1997), which is specially designed for optimal matching analysis but can only deal with a limited number of sequences (up to 150 at a time). The other, Transition Data Analysis (TDA) is now widely used in statistics and data analysis. TDA offers a large number of functions with which to describe complex sequences as well as for the comparison of sequences (allowing for multiple sequences for each individual). Cluster analysis and correspondence analysis have also been included in the more recent versions.

Notes

1 The following volumes present and compare a number of methods of longitudinal data analysis: Hsiao (1986); Uncles (1988); King (1989); Hagenaars (1990); Magnusson *et al.* (1991); Gilbert (1993); Dale and Davies (1994); Engel and Reinecke (1996); van der Kamp and Bijleveld (1998); Taris (2000); Gershuny and Buck (2001).
2 For details see: http://www.popcouncil.org/hrs/longitudinal/3_0.htm
3 For details see: http://www.statsoftinc.com/textbook/sttimser.html#1general http://www.statsoftinc.com/textbook/sttimser.html#systematic
4 For details see: http://www2.chass.ncsu.edu/garson/pa765/time.htm
5 See: http://www.astro.psu.edu/statcodes/sc_timeser.html http://www.decisioneering.com/cbpredictor/
6 See: http://www2.chass.ncsu.edu/garson/pa765/structur.htm
7 http://www2.chass.ncsu.edu/garson/pa765/logit.htm
8 For details see: http://www2.chass.ncsu.edu/garson/pa765/logit.htm http://www.jr2.ox.ac.uk/Bandolier/band25/b25–6.html
9 For details see: http://multilevel.ioe.ac.uk/index.html
10 See: http://multilevel.ioe.ac.uk/index.html

Conclusions

The use of retrospective and prospective longitudinal data ensures a more complete approach to social empirical research. With such data, social investigators have powerful instruments to get to the heart of many processes of social change and to craft effective policies for addressing social problems. Dynamic data are the necessary empirical basis for a new type of dynamic thinking. Sociologists have a long-standing problem in understanding the relationship between social structure and individual behaviour: people's actions are both constrained and enabled by social structures and social norms, which 'impose order and restrictions' on life-courses. But social structures are themselves constituted by aggregations of individual behaviour (Mayer, 1991; Elder, 1992; Gershuny, 1998, 2000). Indeed, macro-changes in both social and economic structures affect individuals and households, producing and interacting with changes at the micro-level: the main objective of longitudinal analysis is indeed to provide both social scientists and policy makers with micro-data to improve our understanding of the incidence, pattern, duration of such processes of change and of their impact on people's everyday life (Rose, 1999: 4, 7).

However, while on the one hand such data can, potentially, provide fuller information about individual behaviour, on the other hand, the use of longitudinal data does pose crucial theoretical and methodological problems. This is one of the reasons why, although longitudinal data is increasingly available, social science research still tends to restrict itself to cross-sectional analyses. Other reasons are: dynamic analysis is, in itself, highly complex; longitudinal studies are usually very expensive both in terms of the money and of the time and energy they require (not only must it be ensured that the same subjects can be measured repeatedly over the course of many years, great risks are also run if the research team cannot be preserved over the duration of the study) and, last but not least, the world of longitudinal research is extremely heterogeneous.

So what guidelines can be offered to readers to help them find their way through the labyrinth? Some useful, general hints can be found in the literature (Menard, 1991):

- If a study is not interested in measuring change, if there is no interest in causal relationships, or, if causal and temporal order are known, then cross-sectional data and analysis may be enough. When conducting cross-sectional research, at most the correlation between variables can be assessed and thus whether variables co-vary ascertained (van der Kamp and Bijleveld, 1998). However, repeated cross-sectional designs may be appropriate if it is thought that the problem of panel conditioning may arise, as a result of repeated interviewing or observation, in a prospective panel.
- If a study aims to investigate historical change – changes over time – then longitudinal data is indispensable, as the only way to investigate change is by collecting repeated measurements. In the social sciences, dynamic data must be available when estimating the parameters of each process.
- If change is going to be measured over a long timespan, then a prospective panel is the most appropriate design for the study, because independent samples may differ from one another unless both formal and informal procedures for sampling and data collection are rigidly replicated for each wave of data gathered. Indeed, it is important to remember that a period of time must elapse before any analysis of social change can be effected, and, long-term in-depth analyses require data gathered from a considerable number of waves.
- If change is to be measured only over a relatively short time (weeks or months), a retrospective design may be appropriate for data concerning events or behaviour, but probably not for attitudes or beliefs.
- In order to combine the strengths of panel designs and the virtues of retrospective studies, a mixed design employing a follow-up and a follow-back strategy seems appropriate (Blossfeld and Rohwer, 1995).

A cross-sectional study may be sufficient if the research problem does not require a dynamic approach. However, if the research hypothesis does demand a dynamic approach, then it is worth investing more time and money and setting up a longitudinal study: the costs of longitudinal designs pay off in terms of the appropriateness with which certain research questions can be addressed.

To encourage greater use of longitudinal data, there must be more exchanges of information between scientists and researchers: those who already perform longitudinal research, those who are approaching it, and those who would like to use it but do not know how. For most researchers,

longitudinal research is still an unexplored land: fascinating but dangerous. Some key reasons for encouraging the spread of longitudinal social research are: production of high quality data; accessibility of data and training (Ghellini and Trivellato, 1996). Above all, any data produced must be high quality data; this is closely linked to the procedures used to develop both the study and the process used to evaluate the data produced. In this latter context, information distribution is an important investment activity which should be encouraged, as it offers a way of obtaining feed-back which will, in its turn, help improve the study. There must be a clear policy about how files are to be made available for public use, a policy which guarantees that all the data remain confidential and which satisfies the growing needs of research. It is hardly necessary to stress the importance of such a policy for panel data on households: these panels are usually launched precisely because they (regularly) offer an opportunity to analyse social change at the micro level.

Appendix 1

List of longitudinal studies mentioned in the book

Belgium

Belgian Socio-Economic Panel (SEP)

Type: Panel study launched in 1985
Original sample: 6,471 households
Purpose: to analyse income distribution, poverty and the effectiveness of the Belgian social security system.

Panel Study of Belgian Households (PSBH)

Type: Panel study started in 1992
Original sample: 4,439 households
Purpose: to collect information about household change, education, occupation, employment, income, expenses, wealth, health, social activities, time-spending, values, relations, role patterns, housing, migration and mobility.

Canada

Survey of Labour and Income Dynamics (SLID)

Type: Rotating panel study started in 1993
Original sample: 15,000 households
Purpose: to provide national data on the fluctuations in income that a typical family or individual experiences over time, allowing insight into the nature and extent of poverty in Canada.

Denmark

The IDA Database for Labour Market Research

Type: Linked panel that contains annual information covering the period 1980–98
Original sample: the database collects information on all persons in the population and all establishments with paid employees.
Purpose: to provide access to coherent data about persons and establishments.

European Community

European Community Household Panel (ECHP)

Type: Panel study started in 1994
Original sample: 60,819 households
Purpose: to investigate, at European Community level, both poverty and social exclusion.

France

Socio-Economic Survey of Lorraine – Panel des Ménages Lorrains (ESEML)

Type: Regional panel study started in 1985 and ended in 1990
Original sample: 2,092 households (the first wave in 1985 was limited to a subsample of 715 households)
Purpose: to collect information about household composition and personal demographic characteristics; housing; income, education; employment/unemployment, poverty and life events.

Germany

German Life History Study (GLHS)

Type: Retrospective cohort study started in 1981–83
Original sample: six different birth cohorts for West Germany; four birth cohorts in East Germany
Purpose: to collect data about life events and about the more important activities of subjects (duration and frequency).

German Socio-Economic Panel (GSOEP)

Type: Panel study launched in 1984
Original sample: 5,921 households

Purpose: to monitor household change; occupational and family biographies; employment and professional mobility; earnings; health; personal satisfaction.

Great Britain

National Child Development Study (NCDS)

Type: Cohort study started in 1958
Original sample: 17,414 individuals
Purpose: to improve understanding of the factors affecting human development over the whole life span.

British Cohort Study (BCS70)

Type: Cohort study started in 1970
Original sample: 17,198 individuals
Purpose: to monitor physical, educational, social and economic development.

ONS Longitudinal Study (LS)

Type: Linked panel started in the early 1970s (the original sample was selected from 1971 Census)
Original sample: 1 per cent of the population of England and Wales (approximately 500,000 individuals)
Purpose: to collect data on vital events: live and still births to women, cancer, deaths.

Women and Employment Survey (WES)

Type: Retrospective study carried out in 1988
Original sample: 5,588 women in Great Britain aged 16–59 and the husbands of 799 of the married women.
Purpose: to establish what factors determine whether or not women are in paid work and to identify the degree to which domestic factors shape women's lifetime labour market involvement.

British Household Panel Study (BHPS)

Type: Panel study launched in 1991
Original sample: 5,511 households
Purpose: to further the understanding of social and economic change at the individual and household level in Britain.

Hungary

Hungarian Household Panel Study (HHP)

Type: Panel study launched in 1992
Original sample: 2,611 households
Purpose: The HHP focuses on dynamic changes in the labour market, income inequalities, the life prospects of the various strata of the population, and the financial and economic strategies of households.

Italy

Bank of Italy Survey of Household Income and Wealth (SHIW)

Type: Cross-sectional study started in 1965; in 1989 a panel section was introduced
Original sample: in 1989 about 15 per cent of the sample (1,208 households) was obtained by re-interviewing families already interviewed in 1987
Purpose: to gather information concerning the economic behaviour of Italian families at the microeconomic level.

Longitudinal Study of Italian Families – Indagine Longitudinale sulle Famiglie Italiane (ILFI)

Type: Panel study started in 1997 with a first, retrospective wave (in 1997). The second wave (1999) has just finished, while the third wave (2001) is, currently, being launched.
Original sample: 4,714 households
Purpose: to collect information on a sample of Italian families (family composition, income sources and levels, demographic and social characteristics) and to study social change.

Ireland

Irish Panel Study, now Living in Ireland Panel Survey (LII)

Type: Panel study started in 1987; in 1994 it became the Irish component of the ECHP survey
Original sample: 3,321 households (1987); 4,048 households (1994)
Purpose: to understand Irish living conditions.

Luxembourg

Panel Socio Économique 'Liewen zu Lëtzebuerg/Vivre à Luxembourg' (PSELL)

Type: Panel study started in 1985
Original sample: 2,012 households
Purpose: to study living conditions of households and individuals in the Grand-Duchy of Luxembourg.

The Netherlands

Socio-Economic Panel Survey (SEP)

Type: Panel study started in 1984
Original sample: 5,000 households
Purpose: description of the main elements of the prosperity of the individual and/or the households and the relationship between the two.

Organisatie voor Strategisch Arbeidsmarktonderzoek (OSA) Labour Supply Panel

Type: Panel study launched in 1985
Original sample: 4,020 households
Purpose: the survey aims to find out about respondents' employment situation, and about their behaviour in the labour market.

Poland

Polish Household Panel (PHP)

Type: Panel study launched in 1987
Original sample: 2,100 households
Purpose: to collect information about household composition and the demographic characteristics of each individual, household incomes, individual incomes, labour force.

Russia

Russian Longitudinal Monitoring Survey (RLMS)

Type: Panel study started in 1992
Original sample: 6,334 households (Round I); 4,718 households (Round II)
Purpose: to measure the effects of Russian reforms on the economic well-being of households and individuals.

Spain

Spanish Household Panel Survey – Encuesta Continua de Presupuestos Familiares (ECPF) or Household Budget Continuous Survey (HBCS)

Type: Rotating panel study started in 1985
Original sample: 3,200 households
Purpose: to collect information on the origin and amount of households' incomes, and the way they are used for consumer spending on specific goods and services.

Sweden

Swedish Level of Living Surveys (LNU)

Type: Panel study started in 1968
Original sample: 6,000 individuals
Purpose: study of health status, working conditions, economic resources, housing standards, family, social integration, education and employment.

Household Market and Non-Market Activities (HUS)

Type: Panel Study started in 1984
Original sample: 2,600 households
Purpose: study of labour market experiences, earnings, schooling, socio-economic background, housing, child care, incomes and taxes, wealth and time use.

Longitudinal Individual Data for Sweden (LINDA)

Type: Linked panel, representative of the Swedish population during 1960 to 1998
Original sample: the database contains information on 300,000 individuals annually
Purpose: to be a complement to surveys such as LNU (The Level of Living Survey) and HUS (The Household Market and Non-market Activities)

Swedish Income Panel (SWIP)

Type: Linked panel set up at the beginning of the 1990s
Original sample: the samples are taken from the register of the total population and from income registers. From the register for 1978 a 1 per cent sample of native born persons (about 77,000 individuals) was taken, as well as a 10 per cent sample of foreign born persons (about 60,000 individuals). A further

10 per cent of the people immigrating each year from 1979 until 1992 was also taken (sample sizes vary between 3,000–7,000 individuals).
Purpose: to study how immigrants assimilate in the Swedish labour market.

Switzerland

Swiss Household Panel 'Vivre en Suisse – Leben in der Schweiz' (SHP)

Type: Multi-purpose panel survey launched in 1999
Original sample: 5,074 households
Purpose: to observe (gross) social change at individual and household level in Switzerland.

United States of America

National Longitudinal Surveys (NLS)

Type: Cohort study started in 1966
Original sample: around 5,000 individuals
Purpose: to gather detailed information about labour market experiences and other aspects of the lives of six cohorts of women and men.

Panel Study of Income Dynamics (PSID)

Type: Panel study launched in 1968
Original sample: 5,000 households
Purpose: the central focus of the data is economic and demographic, with substantial details on income sources and amounts, employment, family composition changes and residential location.

Survey of Income and Program Participation (SIPP)

Type: Rotating panel study started in 1983
Original sample: 26,000 households
Purpose: to measure the effectiveness of existing federal, state and local programs; to estimate future costs and coverage for government programs, such as food stamps; and to provide improved statistics on the distribution of income in the country.

Appendix 2

Longitudinal datasets available in Europe, Russia and North America

This appendix offers the reader a brief overview of the longitudinal datasets used in the book in chronological order. For more detailed information the reader should refer to the books and web pages which deal extensively with the characteristics of these datasets.

The National Child Development Study

The National Child Development Study (NCDS) is a longitudinal birth cohort study of those living in Britain who were born in the week 3–9 March 1958. NCDS was designed to examine the social and obstetric factors associated with stillbirth and death in early infancy among the 17,000 children born in Britain in that one week. To date, there have been six attempts to trace all members of the birth cohort to monitor their physical, educational and social development: one in 1965, when they were aged 7 (NCDS1); one in 1969 (NCDS2), when they were aged 11; one in 1974 (NCDS3), when they were aged 16; one in 1981 (NCDS4), when they were aged 23; and then in 1991 (NCDS5), when they were aged 33. In addition, in 1978, contact was made with the schools they had attended. A sixth sweep was conducted in 1999, and will soon be available for analysis.

The initial sample size was almost 18,000 (17,414) although the number of participants in sweep 5 (1991) was 11,400. Attempts have been made to augment the sample to include additional information and also new immigrants to Britain who were born in the relevant week in 1958. Immigrants were identified from school registers and added to the sample as cohort members at age 11 and 16. A number of specialised follow-up studies have also been carried out, e.g. of people exhibiting respiratory illness symptoms in the 1981 and 1991 surveys. The NCDS is used for a wide range of research, including medical/health research. NCDS also collects information of relevance to investigating women's employment issues – e.g. qualifications, employment, occupation, earnings and income and family composition. Also,

it contains retrospective information on marriage, fertility, employment and housing histories.

The study was initially sponsored by the National Birthday Trust Fund; follow-up studies have been undertaken by the National Children's Bureau and the Social Statistics Research Unit, City University, now known as the Centre for Longitudinal Studies (CLS) and based at the Institute of Education, University of London. Sponsorship for the 1981 and 1991 surveys has come from Government departments and the ESRC and, for 1991, the US National Institute for Child Health and Development.

The data are publicly available through the UK Data Archive at the University of Essex, and on-line at MIMAS (Manchester Infor-Mation and Associated Services) and are well documented for secondary analysis. Access to the data is open to anyone interested, although intending users are asked to commit themselves to ensuring that confidentiality is observed, and to inform the NCDS User Support Group at CLS about their proposed use of the data and any resulting publications. The Data Archive also holds a number of NCDS special sub-studies where additional data has been gathered for samples of cohort members selected for their particular characteristics or circumstances.

Web sites: http://www.cls.ioe.ac.uk/Ncds/nintro.htm
http://www.cls.ioe.ac.uk/Ncds/narchive.htm
http://www.mimas.ac.uk/surveys/ncds/
http://www.mimas.ac.uk/surveys/ncds/ncds_info.html

The National Longitudinal Surveys

The National Longitudinal Surveys (NLS), sponsored and directed by the Bureau of Labor Statistics, US Department of Labor, gather detailed information about the labour market experiences and other aspects of the lives of six groups of men and women. Over the years, a variety of other government agencies, such as the National Institute of Child Health and Human Development, the Department of Education, the Department of Justice, have funded components of the surveys that provide data relevant to their missions. The first set of surveys, initiated in 1966, consisted of four cohorts. These four groups are referred to as the 'older men', 'mature women', 'young men' and 'young women' cohorts of the NLS, and are known collectively as the 'original cohorts'. These cohorts were selected because each faced important labour market decisions, which were of special concern to policy makers. Older men were well into their careers, and were on the threshold of decisions about the timing and extent of their labour force withdrawal. The mature women's cohort was entering middle age and attempting to balance the demands of job and household and childrearing responsibilities.

The cohorts of young men and women were completing their schooling and making initial family and career decisions. Respondents in the mature women's and young women's cohorts continue to be interviewed on a biennial basis, and have been interviewed for over three decades. Both men's cohorts have been retired. The older men's cohort ceased in 1990, with an interview of living respondents and widows or next-of-kin of deceased respondents. Interviews with the young men ceased in 1981. In 1979, a longitudinal study of a cohort of young men and women aged 14 to 22 was begun. This sample of youth was called the National Longitudinal Survey of Youth 1979 (NLSY79). In 1986, a separate survey of all children born to NLSY79 female respondents began, greatly expanding the breadth of child-specific information collected. In addition to all the mother's information from the NLSY79, the child survey includes assessments of each child as well as additional demographic and development information collected from either mother or child. This survey is called the NLSY79 Children. In 1997, the NLS programme was again expanded with a new cohort of young people aged 12 to 16 as of 31 December 1996. This new cohort is the National Longitudinal Survey of Youth 1997 (NLSY97).

Table A2.1 NLS survey plan

Survey group	*Age of cohort in first interview*	*Original sample*	*First/last year*	*No. of surveys*	*No. at last interview*	*Status*
Older men	45–59	5,020	1966/1990	13	2,092[1]	Ended
Mature women	30–44	5,083	1967/1999	19	2,333	Continuing
Young men	14–24	5,225	1966/1981	12	3,398	Ended
Young women	14–24	5,159	1968/1999	20	2,736	Continuing
NLSY79 Youth	14–22	12,686[2]	1979/1998	17	8,399	Continuing
NLSY79 Children	birth–14	[3]	1986/1998	6	4,942	Continuing
Young Adults	15–22	[3]	1994/1998	3	2,143	Continuing
NLSY97 Youth	12–16	8,984	1997/1999	3	8,386[4]	Continuing

Source: http://www.bls.gov/nls/

Notes:

1 Interviews in 1990 were also conducted with 2,206 widows or other next-of-kin of deceased respondents.

2 The sample contains 9,964 respondents eligible for interview.

3 The sizes of the NLSY79 children and young adult samples are dependent on the number of children born to female NLSY79 respondents, which is increasing over time.

4 Fielding of round 3 was begun in October, 1999 and continued through April, 2000. The latest sample size available is from round 2.

The surveys include data about a wide range of events such as schooling and career transitions, marriage and fertility, training investments, child-care usage and drug and alcohol use. The depth and breadth of each survey allow for analysis of an extensive variety of topics such as the transition from school to work, job mobility, youth unemployment, educational attainment and the returns to education, welfare recipiency, the impact of training and retirement decisions.

NLS data files can be ordered via E-mail: http://stats.bls.gov/nlsorder.htm. NLS data are on cohort specific compact discs complete with user-friendly search and retrieval software. This software allows users to search the database for variables, view the codebook information associated with that variable, select and extract variables, and create a codebook unique to the variables chosen.

Web site: http://www.bls.gov/nls/

The Panel Study of Income Dynamics

The Panel Study of Income Dynamics (PSID) is a longitudinal survey of a representative sample of US: individuals (men, women, and children) and the families in which they reside. It has been ongoing since 1968 with a national sample of approximately 5,000 households. Information about the original 1968 sample individuals and their current co-residents (spouses, cohabitors, children and anyone else living with them) is collected each year. Because the original focus of the study was the dynamics of poverty, the 1968 sample included a disproportionately large number of low-income households. To help correct for omissions in representing post-1968 immigrants, a representative national sample of 2,043 Latino households, differentially sampled to provide adequate numbers of Puerto Rican, Mexican-American, and Cuban-Americans, was added to the PSID database in 1990. Information is collected by means of telephone interviewing and, in rare cases where telephone interviewing is problematic, in personal interviews. Information gathered in the survey applies to the circumstances of the family unit as a whole (e.g. type of housing) or to particular persons in the family unit (e.g. age, earnings). While some information is collected about all individuals in the family unit, the greatest level of detail is ascertained for the primary adults heading the family unit.

The PSID provides a wide variety of information about both families and their individual members, plus some information about the areas where they live. The central focus of the data is economic and demographic, with substantial details gathered on income sources and amounts, employment, family composition changes and residential location. Content of a more sociological or psychological nature is also included in some waves of the

Table A2.2 Core topics in the PSID

Income sources and amounts
Poverty status
Public assistance in the form of food or housing
Other financial matters (e.g. taxes, inter-household transfers)
Family structure and demographic measures (e.g. marital events; birth and adoptions; children forming households)
Labour market work (e.g. employment status, work/unemployment/vacation/sick time; occupation, industry; work experience)
Housework time
Housing (e.g. own/rent, house value/rent payment, size)
Geographic mobility (e.g. when and why moved; where head grew up; all states head has lived in)
Socio-economic background (e.g. education, ethnicity, religion, military service, parents' education, occupation, poverty status)
Health (e.g. general health status; disability)

Source: http:/www.isr.umich.edu/src/psid/overview.html

Table A2.3 Major PSID supplemental topics

Housing and neighbourhood characteristics (1968–72, 1977–87)
Achievement motivation (1972)
Child care (1977)
Job training and job acquisition (1978)
Retirement plans (1981–83)
Health: health status, health expenditures, health care of the elderly and parent's health (1986, 1990, 1991, 1993–95)
Kinship: financial situation of parents, time and money help to and from parents (1980, 1988)
Wealth: assets, savings, pension plans, fringe benefits (1984, 1989, 1994)
Education: grade failure, private/public school, extracurricular activities, school detention, special education, Head Start Programs, criminal offense (1995)
Military combat experience (1994)

Source: http:/www.isr.umich.edu/src/psid/overview.html

study. Since 1985, comprehensive retrospective fertility and marriage histories of individuals in the households have been assembled. Other important topics covered by the PSID include housing and food expenditures, housework time and health status. Content of a more sociological or psychological nature is also included in some waves of the study. Beginning in 1985, comprehensive retrospective fertility and marriage histories of individuals in the households have been assembled.

The study is conducted at the Survey Research Centre, Institute for Social Research, University of Michigan (Hill, 1992).

PSID data files are public-use files. Since the start of the study, the PSID data and documentation have been distributed by the InterUniversity

Consortium for Political and Social Research (ICPSR) on magnetic tape and are also available from ICPSR via FTP. Since 1987, the data have also been distributed on CD-ROM. It is also possible to download the PSID data from the PSID homepage: PSID data files, documentation, bibliography, Newsletters, and SAS and SPSS examples for data extraction are also available to users, at no cost, via the Internet.
Web sites: http:/www.umich.edu/~psid/
http:/www.isr.umich.edu/src/psid/overview.html

The Swedish Level of Living Survey

The Swedish Level of Living Survey (LNU), based in the Swedish Institute for Social Research in Stockholm University, started in 1968 with a sample of about 6,000 individuals and involves following up 9,741 cohort members, in the age band 15–75 over four sweeps (1968, 1974, 1981, 1991). Over 7,500 are still participating. The first (1968) Swedish level of living survey was carried out on behalf of the Low Income Committee (Låginkomstutredningen). The second and third level of living surveys were developed in 1974 and 1981, and the institutional basis had, meanwhile, become the Swedish Institute for Social Research (SOFI). The fourth level of living survey was conducted in 1991. All surveys have used paper questionnaires (face-to-face interviews). The population of the survey consists of: 1) persons aged 15–75 years; 2) persons included in the 1974 survey under the age of 76 and still living in Sweden; 3) a new addition of young persons aged 15–21 years; 4) persons immigrating to Sweden 1974–80. The main topics covered are health status, working conditions, economic resources, housing standards, family, social integration, education and employment. All of these studies have potential for secondary analysis, especially in a comparative cross-national framework. Thus the LNU permits comparison with other birth cohort studies.

Level of living data files are stored at the Swedish Social Science Data Service (SSD) as system files for the statistical packages SPSS for Windows. The codebooks for the surveys 1968, 1974, 1981 and 1991 are also available from the SSD; moreover, at the SSD a copy of the 1991 questionnaire in Swedish and in English is available as work documents, as are the interviewer instructions for 1991. The answer sheets from the 1991 survey are also available. Some other documentation is held at SOFI, such as questionnaires in languages other than English, advance letters, etc.
Web sites: http://www.ssd.gu.se/kid/swe/lnu.html
http://www.ssd.gu.se/kid/swe/ssd0719.html
http://www.ssd.gu.se/kid/swe/ssd0720.html

The 1970 British Cohort Study

The 1970 British Cohort Study (BCS70) is a continuing, multi-disciplinary longitudinal study which began when data were collected about the births and families of 17,198 babies born in England, Scotland, Wales and Northern Ireland in the week 5–11 April 1970. At this time, the study was named the British Births Survey (BBS) and it was sponsored by the National Birthday Trust Fund in association with the Royal College of Obstetricians and Gynaecologists. Follow-up studies have been undertaken by the National Children's Bureau and the Social Statistics Research Unit, City University, now known as the Centre for Longitudinal Studies (CLS), Institute of Education, University of London. Since 1970, four more full rounds of data collection have been undertaken: in 1975, 1980, 1986 and 1996. A new survey of the whole cohort was planned for 1999. With each successive attempt, the scope of the enquiry has broadened from a strictly medical focus at birth, to encompass physical and educational development at the age of five, physical, educational and social development at the ages of ten and sixteen, and physical, educational, social and economic development at 26 years (CLS, 1999).

Data have been collected from a number of different sources, and in a variety of ways. In the birth survey (1970), information was collected by means of a questionnaire that was completed by the midwife present at the birth, and supplementary information was obtained from clinical records. The five-year (1975) and 10-year (1980) surveys were carried out by the Department of Child Health, Bristol University and the survey at these times was named the Child Health and Education Study (CHES). In 1975 and 1980, parents of the cohort members were interviewed by health visitors, and information was gathered from head and class teachers (who completed questionnaires), the school health service (which carried out medical examinations on each child), and the subjects themselves (who undertook tests of ability). In both 1975 and 1980, the cohort was augmented by the addition of immigrants to Britain who were born during the target week in 1970. Subjects from Northern Ireland, who had initially been included in the birth survey, were dropped from the study in all subsequent sweeps.

The 16-year (1986) survey was carried out by the International Centre for Child Studies and named Youthscan. In this sweep, 16 separate survey instruments were employed, including parental questionnaires, school class and head teacher questionnaires and medical examinations (including measurement of height, weight and head circumference). The cohort members completed questionnaires, kept two four-day diaries (one for nutrition and one for general activity), and undertook some educational assessments. The most recent 1996 follow-up (BCS70) was

Table A2.4 BCS70 survey plan

Name	*Year*	*Group age of cohort*	*Sample size*
BBS	1970	Birth	17,198*
CHES	1975	5	13,135
CHES	1980	10	14,940
Youthscan	1986	16	11,628
BCS70	1996	26	9,003

Source: http://www.cls.ioe.ac.uk/Bcs70/bintro.htm

Note:
* Achieved Sample – at least one survey instrument partially completed.

carried out by the Social Statistics Research Unit, City University. It was based on a postal survey of those cohort members for whom a current address was available.

Datasets containing the birth, 22-month, 42-month, 5-year, 10-year, and 16-year surveys are now lodged at the UK Data Archive, University of Essex, and on-line at Manchester Infor-Mation and Associated Services (MIMAS). Access to the data is open to anyone interested, although intending users are asked to commit themselves to ensuring that confidentiality is observed, and to inform the Cohort Studies User Support Group at the CLS about their proposed use of the data and any resulting publications. Datasets containing information from the 26-year follow-up, and the 21-year sample survey are currently being prepared at the CLS and will be sent to the Data Archive upon completion.
Web sites: http://www.cls.ioe.ac.uk/Bcs70/bhome.htm
http://www.cls.ioe.ac.uk/Bcs70/bintro.htm

The German Life History Study

The German Life History Study (GLHS) commenced in 1979, promoted by the German Research Society, and continued at the Max Planck Institute of Human Development and Education in Berlin. It is part of a larger research project 'Life-course and Social Change'. The GLHS is now made up of a West German and an East German component (East German Life History Study or EGLHS). The West German Life History Studies (WGLHS) data file contains detailed life-course information for 5,591 men and women of the birth cohorts 1919–21, 1929–31, 1939–41, 1949–51, 1954–56 and 1959–61. These longitudinal data allow analysis of many questions in educational and mobility research, socialisation research, family sociology and migration research to be carried out. In the EGLHS, 2,331 East German women and men (born between 1929–31, 1939–41, 1951–53 and 1959–61) were interviewed between September 1991 and

October 1992. Additionally, 1,265 persons of the initial sample participated in a second questionnaire in summer 1993. Altogether within the GLHS, more than 8,000 life histories covering more than 100 years of German history have been collected.

GLHS data are available to the public and are distributed by the Max Planck Institute of Human Development and Education in Berlin (Centre for Sociology and the Study of the Life-course).
Web sites: http://www.mpib-berlin.mpg.de/en/forschung/bag/index.htm

The Survey of Income and Program Participation

The Survey of Income and Program Participation (SIPP) is a continuous series of national panels with monthly interviewing. The duration of each panel ranges from two-and-a-half to four years. The survey uses a four-month recall period, with approximately the same number of interviews being conducted in each month of the four-month period for each wave. Interviewing for the first panel, the 1984 panel, began in October 1983 with a sample of approximately 26,000 designated households. For the 1984–93 panels, a new panel of households was introduced in February of each year. With the 1996 panel the SIPP questionnaire was redesigned, and a new sample design was introduced. This new four-year panel consisted of 36,700 sample units (households): households are to be interviewed 12 times from April 1996 through March 2000.

The SIPP sample is a multistage-stratified sample of the US civilian non-institutionalised population. Sample size ranges from approximately 14,000 to 36,700 interviewed households. Interviews are conducted by personal visit and by decentralised telephone: the 1996 panel SIPP interviews were conducted using a computer-assisted interview on a laptop computer. The primary survey document is the questionnaire. In each wave a separate questionnaire is completed for every person 15 years old and over living with original sample members.

The SIPP content is built around a 'core' of labour force, program participation and income questions designed to measure the economic situation of persons in the United States. The survey has also been designed to provide a broader context for analysis by adding questions on a variety of topics that are not covered in the core section. These 'topic modules' are assigned to particular interviewing waves of the survey. Topics covered by the modules include personal history, child care, wealth, program eligibility, child support, disability, school enrolment, taxes and annual income.

The SIPP was originally sponsored by the Census Bureau and the Department of Health and Human Services (HHS). Work was well under way for a February 1982 start of the survey when HHS had to withdraw its support

due to funding problems. As a result, the survey was postponed until the Census Bureau received adequate funding from Congress to conduct the survey.

Data are periodically released in cross-sectional, topic module and longitudinal reports. The SIPP team also releases public use files containing the core data on income recipiency and program participation. These files are available currently for all waves; longitudinal files are also available.
Web sites: http://www.sipp.census.gov/sipp/sippov98.htm
http://www.sipp.census.gov/sipp/sipphome.htm

The German Socio-Economic Panel

The German Socio-Economic Panel (GSOEP) is a wide-ranging representative longitudinal study of private households in Germany. In 1984, 5,921 households containing 12,245 people participated in the 'GSOEP West', 1,400 of which were headed by non-Germans: they constituted a separate sample of the immigrant component in the West German population, which immigrated in the 1960s and early 1970s. As early as June 1990, i.e. before currency, economic and social union, the survey was extended to include the territory of the former German Democratic Republic (GDR): 2,179 households with 4,453 people were surveyed in the GDR. This sample constituted the 'GSOEP East' sample. In 1994–95 a new immigrant sample was introduced. The 1998 wave of the data includes 4,285 households with 8,145 people for the GSOEP West sample, and 3,730 people in 1,816 households in the GSOEP East sample. In 1998 – for the first time after 15 years – the GSOEP was extended by a supplementary sample with 1,957 people in 1,079 households. This new sample was added to: ensure stability of the case numbers and to permit analysis of panel effects and survey non-response.

Thus, there are five subsamples; each of these was drawn in a different multi-step random sampling process:

1 Subsamples A and B (started in 1984) cover the old Federal Republic (prior to unification).
2 Subsample B (started in 1984) was deliberately intended to over-sample each of five main nationalities of foreigners (Turkey, Greece, Yugoslavia, Spain and Italy).
3 Subsample C (started in 1990) represents the former GDR.
4 Subsample D (started in 1994–95) includes people living in private households in the western states of Germany in 1994 or 1995 and containing at least one household member who has moved from abroad to Germany after 1984. It is divided into two different subsamples: subsample D1 with 236 households and subsample D2 with 295 households.

Table A2.5 GSOEP dataset: starting sample size in Wave 1

Sample	*Year*	*Households (net)*	*Persons (gross)*	*Respondents (net)*	*Children*
100%					
A and B	1984	5,921	16,205	12,245	3,915
C	1990	2,179	6,131	4,453	1,591
D1	1994	236	733	472	248
D2	1995	295	915	622	283
95%					
A and B	1984	5,624	15,397	11,610	3,711
C	1990	2,071	5,818	4,229	1,510

Source: Frick, 1998 (GSOEP documentation, distributed to users on CD-ROM).

5 Subsample E (started in 1998) is a random, 'refresher' sample covering all existing subsamples.

All members of the households aged 16 or older are questioned once a year. Respondents who move, continue to take part in the study as long as the move is within the Federal Republic of Germany (prior to reunification this did not include the GDR).

The data supply information about both objective and subjective living conditions, about the process of change in various areas of life and about the links between these areas and the changes themselves. Indeed, the GSOEP covers a wide range of subjects including: household composition; occupational and family biographies; employment and professional mobility; earnings; health and personal satisfaction as well as subjects covered in the topic modules of the survey. These modules cover such topics as: social security; education and training; allocation of time; family and social services.

The GSOEP was founded as a project of the Special Research Area 3 'Microanalytical Basis of Social Politics' at the universities of Frankfurt (Main) and Mannheim. It is independently funded through the Deutsche Forschungsgemeinschaft/German National Science Foundation (DFG) and based at the German Institute for Economic Research (DIW) in Berlin. The Center for Policy Research at Syracuse University, in co-operation with the DIW, has prepared an English language public-use version of the GSOEP for use by the international research community which offers a 95 per cent random sample of the original data. The public use file of the GSOEP, with anonymous micro data, is provided free of charge to universities and research centres. Use of the data is subject to special regulations. To obtain the GSOEP data, the potential user first has to sign a data transfer contract with the DIW. Once the contract has been signed, the user will receive the data. GSOEP data are disseminated in several formats, on CD-ROM. The formats

Table A2.6 Special topical modules, GSOEP dataset

Wave and sample	*Description*
1985 Wave 2 (West German residents and foreigners)	Marriage and family biography (retrospective questions)
1986 Wave 3 (West German residents and foreigners)	Social origins, first job (retrospective residents questions), neighbourhood
1987 Wave 4 (West German residents and foreigners)	Social security, early retirement, persons requiring care and child care
1988 Wave 5 (West German residents and foreigners)	Assets
1989 Wave 6 (West German residents and foreigners)	Further education or training and residents and qualifications
1990 Wave 7 (West German residents and foreigners)	Use of time and preferences
1990 Wave 1 (East Germans)	Base questions (labour market + subjective indicators)
1991 Wave 8 (West German residents and foreigners)	Family and social services
1991 Wave 2 (East Germans)	Family and social services (shortened version plus repetition of subjective indicators and labour market indicators of Wave 1 base questions)
1992 Wave 9 (West German residents and foreigners)	Social security and poverty (partly a repetition of Wave 4)
1992 Wave 3 (East Germans)	Social security and poverty (partly a repetition of wave 4), labour market indicators and biographical information (retrospective questions)
1993 Wave 10 (West German residents and foreigners)	Further education or training (shortened repetition of Wave 6)
1993 Wave 4 (East Germans)	Further education or training, labour market
1994 Wave 11/5	Neighbourhood, values, and expectations
1994 Wave 1 (Immigrants, subsample 1)	Same as Wave 11/5 plus immigration history and biography
1995 Wave 12/6	Partial repetition of Wave 7 – use of time and preferences, increased range of income questions
1995 Wave 1 (Immigrants, subsample 2)	Same as Wave 12/6 plus immigration history and biography
1996 Wave 13/7	Repetition of social network questions (Wave 8/2)
1997 Wave 14/8	Social security and poverty (repetition of Wave 9/3)
1998 Wave 15/9	Ecology and environmental behaviour (indirect taxation)
1999 Wave 16/10	Expectations, use of time
2000 Wave 17/11	Further education or training, labour market

Source: Frick, 1998 (GSOEP documentation, distributed to users on CD-ROM).

include SAS, STATA, SPSS, ASCII and TDA. Training workshops for GSOEP users are held annually inside and outside Germany.
Web sites: http://www.diw.de/english/sop/index.html
http://www.diw.de/english/sop/uebersicht/

Household Market and Non-market Activities

The Swedish project Household Market and Non-market Activities (HUS) started in 1980. In 1984 the first main survey was carried out, a comprehensive interview survey that was followed by smaller surveys in 1986, 1988, 1991, 1993, 1996 and 1998. Refresher samples have been added to the panel in 1986, 1993, 1996 and 1998. Data cover many topics the most important being: labour market experiences, current employment, earnings, schooling, socio-economic background, housing, child care, incomes and taxes, wealth and time-use. Event history data are available for labour market events, household changes, child care and housing.

The 1984 survey was based on a random sample of about 2,600 households. This sample excluded people 75 years or older, those who lived in institutions or abroad and those who did not speak Swedish well enough for an interview. In households with two spouses both spouses were interviewed. In some households a third adult was interviewed too. Until 1998 data from all first-time respondents were collected in face-to-face interviews using paper and pencil questionnaires. Data from panel members have always been collected in computer-assisted telephone interviews (CATI). In 1998 all interviews were done by telephone.

In 1986, the 1984 sample was interviewed once more: this time a telephone interview was conducted to obtain information on changes in family composition, housing, employment, wages and child care. As a complement to the panel, a new supplementary sample of households was interviewed. The supplement consisted partly of the members of the 1984 households who were over 18 or who had moved in with someone included in the 1984 sample, and partly of a new random sample of some 800 households. The individuals included in the supplement were asked approximately the same questions as in the 1984 personal interview.

The 1988 survey was considerably smaller than the previous ones: it was addressed exclusively to participants in the 1986 survey, and consisted of a self-enumerated questionnaire with a non-respondent follow-up by telephone.

In 1991, another self-enumerated questionnaire was administered to the panel. An attempt was also made to include the new household members who had moved into sample households since 1986 in the survey, as well as the young people who had turned 18 after the 1986 survey.

Table A2.7 Effective HUS sample size (net of non-response) by wave and sample

Wave	*Sample*	*Number of individuals*
1984		2,619
1986	Panel	1,949
1986	Refresher	1,014
1988	Panel	2,297
1991	Panel	2,052
1993	Panel	1,811
1993	Refresher	1,643
1993	Nonresponse	733
1996	Panel	2,963
1996	Refresher	276
1998	Panel	2,347
1998	Refresher	1,565

Source: http://www.isr.umich.edu/src/psid/inventory_table_links/swedish_overview.do.htm

As regards its design and question wording, the 1993 survey was a new version of the 1986 survey. The 1993 survey was made up of four parts: 1) the panel survey, which was addressed mainly to respondents in the 1991 survey, with certain additions; 2) the supplementary survey, which focused on a new random sample of individuals; 3) the non-response survey, which encompassed respondents who had participated in at least one of the earlier surveys but had since dropped out; and 4) the time-use survey, which included the same sample of respondents as those in the panel and supplementary surveys (Klevmarken and Olovsson, 1993; Flood, Klevmarken and Olovsson, 1997). Time-use interviews were done in 1984 and in 1993.

HUS data can only be used for academic research and they are only available for this purpose in anonymous form. Each user has to sign a contract stipulating that data will only be used for research and that the user will not publish or otherwise make public data for single individuals or households or try to find out the identities of the respondents. A general description of the HUS surveys, code books, test dataset and instructions on how to obtain access to data are on the Internet site: http://www.handels.gu.se/econ/econometrics/hus/husin.htm

From this site, datasets are distributed either as zip-files attached to an E-mail message or on diskettes, by regular mail. HUS data can also be obtained from the Swedish Social Science Data Service (SSD), Göteborg University through its Internet home page: www.ssd.gu.se. Data and code-books are distributed on a CD. Normally, HUS data are distributed as SAS-files. The latest files distributed from the SSD are in a more general format (ASIDE) which can be read by all computers. The details of the surveys have been documented in a set of code-books. Interviewing has been done in Swedish and there is a Swedish code-book for each wave and sample. They have not

been printed but are available as Word documents. Translations into English are currently available for waves 1984–96 as Word-files. The translation for the 1998 wave will probably be ready for distribution in the year 2001. For the period 1984–93 there are also printed code-books in English. All data files and documentation can be obtained at a service charge of approximately 500 USD.

Web sites: http://cent.hgus.gu.se/econ/econometrics/hus/husin.htm
http://cent.hgus.gu.se/econ/econometrics/hus/order/husorder.htm
http://www.ssd.gu.se/enghome.html

The Socio-Economic Panel Survey

The purpose of the Dutch Socio-Economic Panel Survey (SEP) is to describe the main elements of the prosperity of the individual and/or the households and the relationship between the two. The elements concerned are: transfers of income, living conditions, saving and consumption patterns, hours worked, domestic production and an evaluation and perception of prosperity. The idea of carrying out a socio-economic panel survey was conceived by CBS (Statistics Netherlands) in 1977. A pilot study was carried out in September 1983. The survey commenced in April 1984 based on a sample of Dutch households, covering approximately 5,000 households (11,809 individuals): the questionnaire was administered to every adult member of the household (aged 15 or over). All persons who participated in one or more waves were included in the next wave with the exception of those who had left the population (by death, emigration, entering an institution) or who had refused to participate again. The April 1984 SEP sample was a two-stage address sample. All the households living at one address (with a maximum of three) were included in the panel: persons in detention, institutions and in homes for the aged or infirm were not included in the sample. During the 1984–89 period two waves per year were carried out, in April and October of each year. In 1990 this was changed to one wave per year: the two questionnaires have been combined and are conducted annually in April. In view of the high non-response to the first wave (approximately 48 per cent) and of the attrition rate, additional addresses had to be recruited in both 1985 and 1986. However, no new addresses were added for the SEP waves started in April 1987, April 1988, April 1989 and April 1990. An extra 570 addresses were added for October 1987, 400 for October 1988 and for October 1989 slightly less than 400 additional addresses were required (Lemmens, 1991).

Dutch SEP data can be ordered via CBS-Statistics Netherlands (Division Sociaal-Economische Statistieken). English translations of the documentation and the variable/value labels of a number of waves of the Dutch SEP are now available through the CentER Institute, Tilburg University, Faculty of

Economics and Business Administration. Researchers and others who intend to use the data should contact CentER to receive a password in order to download the translations.

The following documentation is currently available in English:

- 1988: Complete documentation of the waves of April 1988 and October 1988. This includes an introduction to the panel setup and the survey, as well as an explanation of the derived variables.
- 1992: Complete documentation of the wave of May 1992 consisting of an introduction to the panel setup and the survey; a questionnaire and documentation on variables used by the Socio-Economic panel 1992; an explanation of the derived variables.
- 1994: Complete documentation of the wave of May 1994.
- 1984–95: Complete documentation of the longitudinal dataset 1984–95.

Web sites: http://www.cbs.nl/en/
http://center.kub.nl/research/facilities/sep.html

Panel Socio-Economique 'Liewen zu Lëtzebuerg/ Vivre à Luxembourg'

The Panel Socio-Economique 'Liewen zu Lëtzebuerg/Vivre à Luxembourg' (PSELL) is a longitudinal study on living conditions of households and individuals in the Grand-Duchy of Luxembourg. PSELL I (1985–94) was launched in 1985, with a sample of 6,110 individuals living in 2,012 households. PSELL II, started in 1994 and is based on a representative sample of 2,978 households and 8,232 individuals. Information is collected by means of face-to-face interviewing. The initial sample was a simple random sample of persons drawn from a register from the Inspectorate General for Social Security. The basic sample represents 97 per cent of the population living in the country. Excluded are: 1) foreign residents who have no links with the country's social security system or who do not live in a household where at least one of the members has such links; 2) elderly persons living in a collective household such as an old people's home. In 1991 an extension was added to the sample. These households had been already selected in wave one, but were not included in the sample at the time: in 1991 these households and their split-offs were included.

Unlike other longitudinal prospective studies which gather data at two levels (individuals and households), in the PSELL there are three distinct data collection units adopted in the survey, namely:

- households;
- income groups within a household;

- individuals.

A household consists of all persons who live together in a dwelling unit (house, apartment, group of rooms or single room). Persons within a household may, or may not, be related to each other. Income Groups are defined as groups of persons within a household, who constitute an economic unit: in a household in which several persons have individual income, different economic arrangements are possible.

The PSELL is carried out by CEPS/INSTEAD (Centre d'Études de Population, de Pauvreté et de Politiques Socio-Économiques, the International Networks for Studies in Technology, Environment, Alternatives, Development). Data are accessible in Luxembourg on a mainframe.
Web sites: http://www.ceps.lu/psell/pselpres.htm
http://www.ceps.lu/projects.htm#PSELL

Enquête Socio-Économique auprès des Ménages Lorrains – Panel des Ménages Lorrains

The Enquête Socio-Économique auprès des Ménages Lorrains – Panel des Ménages Lorrains (ESEML) was launched in 1985 jointly by Équipe de récherche en Analyse Dynamique des Effets des Politiques Sociales (ADEPS, University of Nancy II) and by the Direction Régionale en Lorraine de l'Institut National de la Statistique et des Études Économiques (INSEE). The initial sample size was 2,092 households, although the first wave in 1985 was limited to a sub-sample of about 700 households. The data cover the years 1985–90: the study ended in 1990.

The reference population was anyone living in Lorraine, except persons living in a collective household (e.g. in an old people's home). The original sample was a simple random sample of persons drawn from the Echantillon Démographique Permanent (EDP) of INSEE. Each person led to one household. Every person who lived in this household was interviewed and constituted the initial sample (that is, the persons who were followed in the successive waves). In 1988 and 1990 extensions were added to the initial sample by drawing persons born after 1985 into the EDP.

Standard topics in the ESEML are: household composition and personal demographic characteristics; housing; incomes (on a monthly basis), education; employment/unemployment; biography (education, employment, family background); and life events. Special topics covered by single waves are: housing background; subjective indicators (poverty), difficulty in paying some expenditures; economic behaviour after a large decrease in income, beneficiary of the Guaranteed Minimum Income; project to create a self-employed activity; non-monetary incomes; household assets; duration and

cost of nursing; services granted to elderly persons; debts, intra-family monetary transfers.

The six available waves of data are accessible in the laboratory ADEPS (Nancy) on Unix station or on PC after transfer of the data. Accessibility is decided individually. The data are completely anonymised (no name and no residence code).

Web site: http://www.ceps.lu/paco/pacofrpa.htm

The Belgian Socio-Economic Panel

The main purpose of the survey is to analyse income distribution, poverty and the effectiveness of the Belgian social security system. The administrative unit responsible for the survey is the Centre for Social Policy (CSP), University of Antwerp (UFSIA). The population of the survey consists of all private households resident in Belgium. It includes resident foreigners, and excludes people in institutions, as well as persons without permanent addresses. It is estimated that the survey population covers more than 98 per cent of the total Belgian population. At the moment, three waves of the Belgian panel study are available (1985, 1988, 1992).

The 1985 wave of the data includes 6,471 households; 3,800 in 1988 and 3,800 in 1992 (which included a new sample of 900 households). Sampling took place in two stages: first a number of municipalities were selected, second, within each municipality, a number of households were selected. The first wave (1985) was conceived as a cross-sectional survey. The interviewing was done by an external commercial research organisation, it began in May 1985 and ended in May 1986 and was wholly administered through personal visits. In 1988 the survey was extended into a panel survey and administered through a mixture of personal interviews and mail questionnaires. Mail questionnaires were sent to all households, except the very old (head 75+ years) and households of which the head had only primary education. Households, who did not qualify for a mail questionnaire (about one-third of the sample), as well as households who did not respond to it, were approached for a personal interview. Mail questionnaires were administered in the following way: 1st week: letter announcing the questionnaire; 2nd week: questionnaire with accompanying letter; 3rd week: reminder (printed out); 5th week: 2nd questionnaire. In principle, all members of wave one households were followed up for the second wave, regardless of their family status in the first wave. Students attending universities were considered to be still part of their original household. This applied also to people who had gone into institutions such as prisons and hospitals, if this was for a relatively short period. In the case of people moving to another town, the interview was

assigned to another interviewer, who lived nearer. However, people who entered the population between waves 1 and 2, and who did not live in the same household as a wave one sample member, had no chance of being included in the wave two sample. To achieve a larger sample new households were added to the original panel sample in the 1992 survey. These additional households were obtained via a new sample which had a design identical to the original panel sample. The same survey procedure as in 1988 was employed. Interviewing for the third wave started in December 1991 and ended in March 1992. It was administered through a personal visit by the interviewer after the interview had been announced by means of an introductory letter. Respondents were offered the chance to fill in the questionnaire themselves, in which case the interviewer only collected the interview and carefully checked to see that it had been filled in correctly (Cantillon, 1990; Deleeck *et al.* 1992; Delhausse, 1992).

Belgian SEP data (SPSSx, SAS or ASCII files) are available via Data Archives. SEP data can also be used via the Luxembourg Income Study (LIS). Information about data quality and methodological background are available on request.

Web site: http://www.ufsia.ac.be/~csb/eng/septab.htm

The Household Budget Continuous Survey – Encuesta Continua de Presupuestos Familiares

The Spanish Household Budget Continuous Survey (HBCS) – Encuesta Continua de Presupuestos Familiares (ECPF) was launched by the INE (the Instituto Nacional de Estadistica) in January 1985 to provide information about both the origins and the amount of households' incomes, and about the way income is used for a variety of consumption expenditures. The survey is based on 3,200 families. The expenditure on consumption recorded in the survey relates not only to the amount spent on certain goods and services, considered as final consumer goods and services, but also to the perceived value of the goods for self-consumption, self-supply, wages in kind, free or discounted meals and rent imputed to the dwelling in which the household was living.

A methodological change was made to the survey in the third quarter of 1997, with both adjustments in the methods used to gather information and an increase in the size of the sample which allows estimations to be made by autonomous communities. At the same time a new classification of goods and services was introduced under which the different expenditures made by households are coded, to make the information more suitable: i.e. better able, both to meet the needs of National Accounts, and to facilitate international comparisons (especially between European Union members).

Methodological changes in the new project have also required substantial changes being made to the criteria adopted when registering certain expenditure items and to the periods the information refers to. Half of the current sample (over 4,000 households) collaborates for one week per quarter, by keeping a note, in a special notebook, of all the goods and services they have paid for. However, as one week is too short a period to be able to accurately reflect all consumption goods and services the household may acquire, the whole sample (over 8,000 households) is interviewed to obtain information about regular purchases which are made at longer intervals. Every quarter, one-eighth of the sample is replaced, thus every household collaborates for a maximum of eight quarters.

HBCS data are offered in an EXCEL file (which can be browsed using a Microsoft Excel Viewer (Windows 95 version) and in an ACROBAT file (a free-use program that can be captured by selecting Acrobat Reader).
Web sites: http://www.ine.es/welcoing.htm
http://www.ine.es/dacoin/dacoinme/inotecpf.htm
http://www.ine.es/htdocs/dacoin/dacoinci/ecpflsti.htm
http://www.ine.es/dacoin/dacoinci/ecpf/ecpf197i.htm
http://www.ine.es/dacoin/dacoinci/ecpf/ecpf297i.htm

The OSA Labour Supply Panel

The OSA (Organisatie voor Strategisch Arbeidsmarktonderzoek) conducts a survey every two years to collect data about the (potential) labour force in The Netherlands: the OSA Labour Supply Panel. The Supply Panel targets persons between 16 and 65 years of age, who are not in daytime education. The survey aims to find out about respondents' employment situations, and about their behaviour in the labour market. Information is also collected about aspects that may be expected to influence subjects' decision about whether or not to participate in the labour market. The first wave of the OSA Labour Supply Panel was carried out in the spring of 1985. Subsequent surveys have taken place every two years (from 1986 to 1998).

The sample is selected from the total number of households in The Netherlands. All members of the households in the sample that can be regarded as (potential) members of the labour force are interviewed. To guarantee continuity, households that have been involved in previous surveys are eligible for participation in subsequent waves. To limit the decrease in the overall response rate, respondents who are unwilling or unable to take part in future surveys are replaced by newly selected respondents. These are selected on the basis of characteristics of non-responding households.

The questions all respondents have to answer relate to:

Table A2.8 Number of cases in the OSA Labour Supply Panel

Questionnaire	*1985*	*1986*	*1988*	*1990*	*1992*	*1994*	*1996*	*1998*
Participation	4,020	4,115	4,464	4,438	4,536	4,538	4,563	4,780
Year								
1985	4,020	2,755	1,974	1,432	1,072	904	661	505
1986		1,360	1,201	751	584	407	346	263
1988			1,469	988	711	560	438	329
1990				1,267	890	678	476	332
1992					1,279	869	578	388
1994						1,119	754	500
1996							1,310	867
1998								1,596

Source: http://osa.kub.nl/osa_eng/datasets/e6_2_1.html

- personal characteristics, such as gender, date of birth or country of birth;
- family characteristics, marital status, number of children;
- social background, information about the respondent's parents;
- sources of income other than employment, educational background: level and specialisation, period in full-time/part-time education, date of diploma;
- employment background: date, type of job and reason for changing job, and, until the sixth wave, type of and reason for every change of job, number of hours of employment per week (if employed), and income;
- attitudes towards employment.

Moreover, respondents who are currently employed answer questions related to their current job, while respondents who are looking for a job are asked questions which aim to investigate the following aspects: length of time unemployed; job-searching behaviour, frequency and amount of effort; opportunities for finding work; desired and expected salary; type of occupation being applied for.

The OSA Labour Supply database is available for secondary analyses. However, there are a number of access conditions. Information about these conditions can be obtained directly from the OSA.
Web site: http://osa.kub.nl/osa_eng/datasets/e6_2_1.html

The Polish Household Panel

The Polish Household Panel (PHP) started in 1987 with a sample of 2,100 households. It was carried out by the Department of Economics, Warsaw University and sponsored by the Central Statistical Office (CSO). Key topics

of the PHP are: household composition and the demographic characteristics of each individual, household incomes, individual incomes, labour force information. The survey population consists of persons living in private households, excluding police officers, military personnel and members of the 'nomenklatura'. Information is collected by means of face-to-face interviews. The data form part of a cross-sectional household budget survey of the CSO of Poland. Sampling was based on quarterly rotation of households in a yearly cycle and was done once for a four-year period. Two groups of households were surveyed annually. One of them (two-thirds of the sample) remained in the sample for four years, while the families in the other group (one-third of the sample) were replaced every year by new ones. This method made it possible to extract, from the datasets collected for four consecutive years, a subset of households surveyed throughout the whole four-year period. The households in the subset were the candidates for the panel. The data are accessible at the Department of Economics (Warsaw University). Permission to access the data is given individually.
Web sites: http://www.ceps.lu/paco/pacopopa.htm

The Living in Ireland Panel Survey

The Living in Ireland Panel Survey (LII) began in 1994, as the Irish component of the ECHP survey. The survey is carried out annually by the Economic and Social Research Institute (ESRI) in Dublin. In wave 1, there were 4,048 completed sample households containing 14,585 individuals. Of these, 10,418 were eligible for individual interview and 9,904 (95 per cent) were interviewed individually. All individuals in the wave 1 sample were to be followed up in wave 2 and households and individual interviews were to be conducted as long as the person was still living in a private or collective household (that includes boarding or lodging houses and army barracks, but not institutions such as hospitals, nursing homes, convents or prisons) within the EU.

There are two distinct data collection units employed in the survey, namely the household and the individual adult. The household questionnaire is administered to the 'Household Reference Person' or her/his spouse. The individual questionnaire is distributed to each member of the household born in 1977 or earlier. The questionnaire package covers all of the items required for the ECHP, as well as a number of additional items which expand on Eurostat's specifications, such as information on current (as well as previous-year) social welfare and pension receipts (Callan *et al.*, 1996; Watson, 1998). A dedicated questionnaire is administered to collect information on farm size, type of cattle, subsidies, transfers to the farm, etc.
Web site: http://www.esri.ie

Table A2.9 Wave-on-wave response rates in LII

	Wave 1	*Wave 2*	*Wave 3*	*Wave 4*
Households				
Completed households	4,048	3,584	3,174	2,945
Non-response	3,038	794	624	388
Non-sample*	166	97	77	54
Total households	7,252	4,475	3,875	3,387
Household response rate (excluding non-sample)		82%	84%	88%
Individuals				
Number in completed households	14,585	12,649	10,939	10,013
Number in non-response households		2,286	1,781	1,066
Number in non-sample* households		117	219	215
Total individuals		15,052	12,939	11,294
Interviewed	9,904	8,532 (94%)	7,517 (95%)	6,868 (95%)

Source: Watson, 1998.

Note:

* Non-sample households are those where all members deceased, moved to an institution or outside the EU, or households not containing a 'sample person' – someone who was in one of the original households in wave 1.

The Bank of Italy Survey of Household Income and Wealth

The Bank of Italy Survey of Household Income and Wealth (SHIW) started in 1965. Twenty-three further surveys have been conducted since then, yearly until 1987 (except for 1985) and every two years thereafter. The aim of the survey is to gather information concerning the economic behaviour of Italian families at the microeconomic level. The basic survey unit is the household, which is defined in terms of family relationships, that is, as a group of individuals linked by ties of blood, marriage or affection, sharing the same dwelling and pooling all or part of their incomes. Persons living in nursing homes for the aged or ill, in prisons or in military installations are not included.

The survey gathers data on the social and demographic characteristics of household members. Sex, age and relationship to the head of the household are collected for all members; education, professional status and economic sector are recorded for all income recipients. Questions concerning the whole household (family structure, family changes, family incomes and savings, quality and location of the dwelling, family consumption and expenditures, etc.) are answered by the head of the family or by the person most knowledgeable about the family's finances. Questions on individual incomes are answered by each member, unless they are absent. Participation in the survey is voluntary. In order to overcome households' distrust, shortly before

the interviews are scheduled every household is sent a letter explaining the aims of the survey and giving an assurance that all information collected will be treated anonymously. Families are provided, on request, with a copy of the Bank of Italy's publications containing the reports of previous surveys. Nonetheless, refusal to co-operate and 'fear' account for the largest proportion of non-responses.

To allow for better comparisons over time, in 1989 about 15 per cent of the sample (1,208 households) was made up of families who had already been interviewed in 1987. This panel section corresponds to:

- 15 per cent of the households between 1987 and 1989;
- 26.7 per cent between 1989 and 1991;
- 42.9 per cent between 1991 and 1993;
- 44.8 per cent between 1993 and 1995;
- 37.3 per cent between 1995 and 1998;
- 48.4 per cent between 1998 and 2000.

The actual running of the survey is contracted out to a private company, which provides professionally trained interviewers. Data are collected in personal interviews, usually in the first month of a year, about income and savings in the previous calendar year. To reach the planned number of interviews, non-responding families are replaced with other units with similar characteristics.

The sample size was initially set at 3,000 households on the basis of considerations regarding the sampling errors and desired confidence levels. It was raised to 4,000 in 1981 to increase the accuracy of estimates for regional subsamples and to 8,000 in 1986. Sampling took place in two stages, with selection of municipalities in the first stage and families in the second. The sample design was entirely revised in 1986 and made consistent with that used by the Italian National Institute of Statistics (ISTAT) in its Survey of the Labour Force (Brandolini and Cannari, 1994).

The Bank of Italy has now made micro data, gathered between 1977 and 1998, available for users. All these data have been rendered anonymous. The most recent data (1993 to 1998) are distributed almost in their entirety: only information that could lead to the subject being identified indirectly has been excluded. Information that refers to the period 1997–98 is in the historical archive which, however, only contains the subsets of those variables that are considered to be useful for longitudinal analysis. Raw data matrices (ASCII files) can be obtained free of charge together with the basic documentation (questionnaires, methodological notes, list of publications). SAS and STATA instructions are also provided to help load the data. The documentation is in Italian.

Web site: http://www.bancaditalia.it

The British Household Panel Survey

The British Household Panel Survey (BHPS) was launched in 1991. The BHPS is being carried out by the ISER, Institute for Social and Economic Research (incorporating the ESRC Research Centre on Micro-social Change), at the University of Essex. The main objective of the survey is to further the understanding of social and economic change at the individual and household level in Britain, to identify, model and forecast such changes, their causes and their consequences, in relation to a range of socio-economic variables.

It was designed as an annual survey of each adult (16+) member of a nationally representative sample of more than 5,500 households, making a total of approximately 10,200 individual interviews. The same individuals have been re-interviewed in successive waves and, if they have split off from their original households, all adult members of their new households are also interviewed. Children join the sample once they reach the age of 16 (there is also a special survey of 11–15-year-old household members in waves 4 to 5). That is, the sample for the subsequent waves has consisted of all adults in all households containing at least one member who was resident in a household interviewed at wave one, regardless of whether that individual had been interviewed in wave one. Thus, with a few exceptions, an attempt has been made to interview all those individuals in responding households who had refused to participate at wave one, or for any reason had been unable to take part. In addition, a number of households where no contact had been made in wave one were approached for interview in wave two after confirmation that no household moves had taken place between waves. In the 1997 survey, a subsample of 1,000 households from the Great Britain sample for the ECHP survey was added to the BHPS sample and these respondents have been interviewed as part of the BHPS since that time. The numbers of households and individuals from this subsample are not included in Table A2.10.

The BHPS data are deposited in the UK Data Archive within 12 months of the completion of fieldwork. Between the end of fieldwork and the deposit date, the ISER carries out a full programme of data cleaning, missing value imputation and weighting. Data from release nine of the BHPS is now available from the UK Data Archive: it incorporates the core data collected at each wave so far. To obtain access to BHPS data, potential users have to sign a form agreeing to respect the confidentiality of the data they obtain. The data are supplied by the Data Archive free of charge, only the costs of any materials involved (photocopies, diskettes etc.) have to be paid.
Web site: http://www.irc.essex.ac.uk/bhps/

Table A2.10 Number of interviewed households and individual respondents by country* in BHPS (wave 1 to wave 7)

	England		*Scotland*		*Wales*		*Total GB*	
	Households	*Individuals*	*Households*	*Individuals*	*Households*	*Individuals*	*Households*	*Individuals*
1991	4,699	8,774	531	957	281	533	5,511	10,264
1992	4,457	8,406	508	927	260	510	5,225	9,843
1993	4,466	8,215	498	894	268	491	5,232	9,600
1994	4,365	8,099	489	873	273	509	5,127	9,841
1995	4,288	7,915	475	843	270	491	5,033	9,249
1996	4,342	8,134	452	823	269	480	5,063	9,437
1997	4,297	8,064	451	821	276	486	5,024	9,371

Source: http://www.iser.essex.ac.uk/bhps/rwsum.php

Note:
* Includes respondents with full individual interview or proxy interview.

The Panel Study of Belgian Households

The Panel Study of Belgian Households (PSBH) started in 1990 as a project of the 'Impulse program for Social Research' of the Federal Ministry for Science Policy (now called the Federal Department for Scientific, Technical and Cultural Affairs). The project was assigned to the Universities of Antwerp and Liège. In 1992, 4,439 households with over 11,000 members were successfully interviewed. Since then the same persons of the basic sample have been questioned on a yearly basis. Each interview gives about 400 variables on the household level and about 800 variables on an individual level. The topics covered are: demography, composition of the household, education, occupation, employment, income, grants, expenses, wealth, health, social activities, time-spending, values, relations, role patterns, housing, migration and mobility.

In 1993 the PSBH-research team carried out two pilot studies for the ECHP. This European Statistical Bureau project has a similar aim to that of the Belgian Panel Study. At the same time it also facilitates comparative studies in all countries of the EU. Therefore, it was obvious that after making relatively small adjustments to the questionnaire, the PSBH project could be used to provide the Belgian part of the European research project. The 1994-wave was the first one that was part of the ECHP.

Each wave of the PSBH offers accessible information of two different kinds: on the one hand datasets and, on the other, documentation concerning these datasets. The main datasets are available free of charge for all scientific research (including dissertations and doctoral theses) carried out by recognised scientific institutes. Each user has to submit a written application to the promoter of the project. If this formal request is accepted, the database is transferred.

The PSBH data are SAS datasets. Usually transfer to another platform is carried out using an SAS-transportfile (machine-independent format), or in the form of a regular ASCII-file. Physically the transfer can be done via FTP-sessions over the BELNET (network to which nearly all Belgian Universities are connected) or via compressed data on floppies. For reasons of privacy, all details which would make it possible to identify respondents are deleted from the data that is handed over to users. Documentation is available in Dutch. All documentation can be downloaded directly via the Internet in the form of compressed files. If the user would like the documentation in print, on floppy or via email, they can receive it from the PSBH team.

Web sites: http://www.uia.ac.be/psbh/
http://www.sosig.ac.uk/roads/cgi-bin/tempbyhand.pl?query=862917626-11850&database=sosigv3

The Hungarian Household Panel Study

The Hungarian Household Panel Study (HHP) started in 1992 as a joint research project involving the Social Research Informatics Centre (TÁRKI), the Sociology Department of Budapest University of Economics, the Hungarian Central Statistical Office, the National Scientific Research Fund (OTKA) and several other Hungarian institutions. Between 1992 and 1997, a nationwide sample of 2,600 households was surveyed on a yearly basis. The population of the survey consists of all Hungarian non-institutional households.

The HHP focuses on dynamic changes in the labour market, income inequalities, the life prospects of the various strata of the population and the financial and economic strategies of households.

Information is collected by means of three different questionnaires (face-to-face interviews): 1) a household questionnaire (filled in with the help of the most competent member of the household); 2) an individual questionnaire – for each adult in the household (16 years or older); 3) a substitute questionnaire – for each adult not available at the time of the survey (filled in with the help of the most competent member of the household). Each questionnaire contains different blocks. Some of these blocks are wave-specific, others are not.

The original sample used was based on the 1990 census, stratified by county (location), settlement (size), census district (type of urbanisation) and address. The primary sampling unit was the addresses of non-institutional households. A total of 74 settlements and 437 census districts were drawn and, within them, a random sample of 2,000 addresses were selected. An additional sample of the same size was drawn to substitute addresses that

were not available to be part of the sample (unable to answer, moved away, wrong address, dead, etc.). Additionally, a 600 household subsample covers Budapest households, making the total sample also representative of the city of Budapest.

HHP data files, stored as system files for the statistical packages SPSS/PC or as mainframe files, are available in Hungarian and English versions from TÁRKI. There are no restrictions on the scientific use of the data.
Web sites: http://www.tarki.hu/index-e.html
http://www.tarki.hu/common/tarkirol_e.html
http://www.ceps.lu/paco/pacohupa.htm

The Survey of Labour and Income Dynamics

The Canadian Survey of Labour and Income Dynamics (SLID) is a longitudinal household survey conducted by Statistics Canada. SLID is the first Canadian household survey ever to provide national data on the fluctuations in income that a typical family or individual experiences over time, allowing greater insight into the nature and extent of poverty in Canada. Additionally, with the termination of the annual Survey of Consumer Finances, SLID became the source of detailed annual income data starting with calendar year 1998.

The first reference year of the survey was 1993. Starting in 1993, the SLID followed the same respondents for six years: the sample size for panel 1 was approximately 15,000 households and 31,000 adults aged 16 and older. A second panel was introduced in 1996, overlapping the first one for a three-year period. In 1999, panel 3 was introduced and panel 1 was 'retired'. Panel 4 will be launched in 2002. This pattern is being repeated every three years: each panel includes about 15,000 households (approximately 30,000 adults).

A preliminary interview takes place at the beginning of each panel to collect background information. Each of the six years has a split-interview format, with labour topics covered in January and income topics in May. In both cases, questions refer to the previous calendar year. The income interview occurs in May to take advantage of income tax time when respondents are more familiar with their records. In addition, many respondents have given permission to consult their income tax file thus avoiding the income interview.

The data are provided in six files:

- two-year (1993–94) longitudinal person file
- two-year (1993–94) longitudinal job file
- 1994 cross-sectional person file
- 1994 cross-sectional job file

Table A2.11 The sample design of the survey of Labour and Income Dynamics (SLID)

Year	*1993*	*1994*	*1995*	*1996*	*1997*	*1998*	*1999*	*2000*	*2001*	*2001*	*2003*	*2004*
Wave	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Panel 1	•	•	•	•	•	•						
Panel 2				•	•	•	•	•	•			
Panel 3							•	•	•	•	•	•
Panel 4										•	•	•

Source: Cotton and Giles, 1998.

- 1993 cross-sectional person file
- 1993 cross-sectional job file.

The price of each SLID issue is 1,700 Canadian dollars. To maintain confidentiality of respondent information, the microdata released for public use contain somewhat less detail than that which is available on the internal file. The CD-ROM indicates all additional variables that can be used in custom retrievals on the internal file.
Web sites: http://www.ssc.uwo.ca/sociology/longitudinal/Data.htm #Overview of the Survey of Labour and Income Dynamics (SLID) Philip Giles
http://www.statcan.ca/english/IPS/Data/75M0001XCB.htm

The European Community Household Panel

The European Community Household Panel (ECHP) was launched in 1994. It is a source of community and regional level statistical information. Its objective is to supply the European Commission with an instrument for observing and monitoring the standard of living of the population during the process of convergence towards monetary and political union. It presents comparable micro-level (persons/households) data on income, living conditions, housing, health and work in the EU. Although the questionnaire was designed centrally at Eurostat, in close consultation with member states, it allowed enough flexibility to be able to adapt it to national specificities. Thus the ECHP forms the most closely co-ordinated component of the European system of social surveys.

The EC panel study, planned for a total duration of nine years, is conducted in annual cycles. In the first wave (1994) a sample of 60,819 nationally representative households – i.e. approximately 127,000 adults aged 16 years and over – were interviewed in the then 12 member states. The response rate was 71 per cent for the EU as a whole: it varied from 40 per cent in Luxembourg to 90 per cent in Greece and Italy. Austria and Finland have since joined the project (Sweden remains the only exception): the first

wave of the ECHP Austria was launched in September 1995 while the pilot study of the ECHP Finland was conducted in October–November 1996 with an initial sample of 250 households (Eurostat, 1996a). In wave 2, called EU-13, samples totalled some 60,000 households and 129,000 adults: the wave 2 sample was 92 per cent of the wave 1 sample. The 1994–99 waves have been completed and the 2000 wave is in process, although, so far, only the first three waves are available for research purposes. The longitudinal, panel design of the ECHP makes it possible to follow up and interview the same private households and persons over several consecutive years. The members of the initial sample are studied throughout all the cycles of the survey, where new members can join the household and any leavers or all the members of the household are monitored if there is a change of residence within the EU. New household members are interviewed, as long as they belong to a household containing at least one sample person.

The ECHP started in 1991, when Eurostat – the Statistical Office of the European Communities – set up a Task Force on Household Incomes to respond to the strong need felt for information on household and individual income. The Task Force was mandated to assess, together with EU member states, the income data held in registers and in existing national household surveys, and to check whether the available outputs could be satisfactorily harmonised *ex-post*. After the failure of this 'output approach', the decision was taken to launch a specific EU survey (the ECHP), i.e. to adopt an input-oriented approach rather than trying to harmonise existing outputs.

ECHP data are collected by National Data Collection Units (NDUs), either National Statistical Institutes (NSIs) or research centres depending on the country. Dissemination of the database is restricted by means of ECHP research contracts – that are signed with Eurostat and stipulate the strict conditions of data use and access – and are subject to Eurostat's discretion. To meet the increasing demand for ECHP-based statistics and to have direct access to the data, Eurostat decided, together with NDUs, to develop a set of rules that would allow easier direct access to 'anonymised' ECHP micro-data. In November 1997, Eurostat proposed to create a user-friendly and widely documented ECHP Longitudinal Users' Database (UDB) that would meet various 'objective anonymisation criteria'. The first version of UDB was finalised mid-December 1998: the dataset for five sweeps (1994–98) is available on CD-ROM and the idea of producing a more extended version is under discussion. Any request to consult this file must come from an official organisation and access is only permitted after payment of a sum which varies according to the category of each user (Marlier, 1999). Potential users are asked to sign a research contract with Eurostat that covers the assignment to the contractor, on the terms set out in the contract, of the right to use the ECHP users' database, in the form and according to the

Table A2.12 Sample size and changes in the achieved sample size in ECHP*

Country	*Wave 1 Number of households interviewed*	*Wave 2 Number of households interviewed*	*Wave 3 Number of households interviewed*	*Ratio Wave 2/ Wave 1*	*Ratio Wave 3/ Wave 2*
Austria	n.a.	3,382	3,279		97.0
Belgium	4,189	4,012	3,748	95.8	93.4
Denmark	3,482	3,225	2,956	92.6	91.7
France	7,344	6,722	6,542	91.5	97.3
Germany	5,054	4,687	n.a.	92.7	
Greece	5,523	5,219	4,923	94.5	94.3
Ireland	4,048	3,548	3,179	88.5	88.7
Italy	7,115	7,128	n.a.	100.2	
Luxembourg	1,011	962	n.a.	95.1	
The Netherlands	5,187	5,110	n.a.	98.5	
Portugal	4,881	4,916	4,955	100.7	100.8
Spain	7,206	6,521	6,277	90.5	96.3
UK	5,779	4,548	3,420	78.7	75.2
Total (EU 12)	60,819	56,634			
Total (EU 13)	n.a.	60,016			

Source: Eurostat, 1997.

Note:

* In two countries (Italy and Portugal) the Wave 2 achieved sample size exceeded the Wave 1 sample: the formation of new sample households (split-off) exceeded the non-response in these countries. In most cases, the Wave 3 sample corresponded to 90–100 per cent of the Wave 2 sample, with the exceptions of Ireland and the UK.

arrangements specified in the contract. The files contained in the UDB must be used exclusively for research purposes as specified in the contract, excluding, in particular, any possible administrative use. The data may be used by the contractor solely under the conditions and for the purposes described in the contract. The contractor may not process, disseminate or otherwise allow any of the data to be made available or used for any purpose whatsoever other than the research purposes laid down in the contract.
Web sites: http://qb.soc.surrey.ac.uk/surveys/echp/echpintro.htm
http://www.iue.it/LIB/DataSets/LDataSets/echp.htm
http://forum.europa.eu.int/irc/dsis/echpanel/info/data/information.html

It is worth mentioning that there are EU funds available for researchers to visit both the ISER at Essex and the CEPS/INSTEAD in Luxembourg to use UDB data. To visit the ISER – which has played a leading role in analysing the ECHP and is also the UK NDU responsible for collecting the

British component for this survey – researchers should send their application to the European Centre for Analysis in the Social Sciences (ECASS), an interdisciplinary research centre at the University of Essex within the Institute for Social and Economic Research. ECASS is a centre for comparative and longitudinal data analysis, which conducts and facilitates the empirical study of social and economic change by integrating longitudinal and cross-national European datasets, providing the support services required for analyses, and acting as the host for major substantive research programmes. ECASS visitors can collaborate with ISER researchers on its analyses. Researchers interested in UDB data could also visit CEPS/INSTEAD in Luxembourg through bursaries offered by IRISS-C/I (Integrated Research Infrastructure in the Social Sciences at CEPS/INSTEAD). IRISS offers access to the facilities and resources of the institute.
Web sites: http://www.iser.essex.ac.uk/ecass/
http://www.ceps.lu/iriss/iriss.htm

Indagine Longitudinale sulle Famiglie Italiane

The Indagine Longitudinale sulle Famiglie Italiane (ILFI) (Longitudinal Survey of Italian Families) is a prospective panel study that was initially set up by the University of Trento, the Istituto Trentino di Cultura (Trento Institute of Culture) and ISTAT. It is based on a nationwide sample of subjects over 18 years of age. Initially, there were 10,423 subjects who were members of 4,714 families living in 223 Italian municipal areas. The sampling unit is the household. The survey is based on five waves, carried out every two years with a retrospective first wave (in 1997) (Schizzerotto, 1999). The second wave (1999), which was carried out jointly by the Universities of Trento, Milano-Bicocca and Bologna, has just finished while the third wave (2001) is, currently, being launched.

The ILFI seeks to reconstruct the life history of each household member (from birth up to the last wave of interviews – planned for 2005). Information gathered during the first wave dealt with the following situations and events, which had occurred in both the individuals' and the families' lives:

- education and professional training;
- work history;
- household composition and events within the family;
- the family's economic resources;
- episodes of caring and assistance within the family;
- personal health status;
- political attitudes and religious beliefs;
- geographic mobility.

The sections of the questionnaire dealing with education, work and family history are very detailed. This has made it possible to gather precious retrospective information about school and work careers and about the histories of the households included in the sample: housing history, work history of parents, evolution of the family of origin and the origins and history of the family being interviewed. The section on caring also reveals the length of time care and assistance have been, or were, provided within the family. The survey has thus made it possible to collect duration data which are fundamental for any event history analysis.

The second wave (1999) aimed to gather information about the life-course of all the individuals, interviewed in 1997, in the period between that interview and 1999, excluding, of course, any who had died or moved abroad in the intervening two years. Other subjects were added: a) any members of participating families who had had their eighteenth birthday after July 1997; and b) those who were members of families already interviewed in the preceding wave who had left the family of origin to form new families. The second wave paid particular attention to:

1 individual incomes derived from any work done;
2 individual incomes deriving from State emoluments;
3 individual incomes supplied by family or friends;
4 access to nursery and educational services for infants and young children;
5 access to health services;
6 access to caring services for the disabled and the elderly;
7 the strength of the family and friendship network offering both material and non-material assistance.

Last, the third wave (2001), intends to examine two specific themes involving the sphere of work. First, a study of the transition from the school system to the labour market, which will be carried out by constructing a life history calendar, which gives detailed information that will make it possible to accurately identify the events that take place between leaving school and starting to look for a first job. Second, the processes involved in career mobility are to be studied in depth by means of a series of questions which are designed to identify, very accurately, the reasons why employees lose, or have lost, their jobs (sacking/redundancy, early retirement, the decision to hand in notice voluntarily, or 'forced' resignation, etc.).

ILFI data can be ordered and obtained after payment of a sum to cover costs. Access to ILFI data is restricted to users with a specific contract: this contract has to be signed not only by the person who is responsible for the research group but also by each member of the research team. Once the contract has been signed, the potential user(s) will receive the data.

For information, contact Prof. Schizzerotto, the director of the ILFI survey, at the following E-mail address: antonio.schizzerotto@unimib.it.
Web site: http://www.unitn.it/unitn/numero16/indagine.html

The Russian Longitudinal Monitoring Survey

The Russian Longitudinal Monitoring Survey (RLMS) is a household-based survey designed to measure the effects of Russian reforms on the economic well-being of both households and individuals. These effects are measured by a variety of means: detailed monitoring of individuals' health status and dietary intake; measurement of household-level expenditures and service utilisation; and collection of relevant community-level data, including region-specific prices and community infrastructure data. The RLMS survey instruments were designed by an interdisciplinary group of Russian and American social science and biomedical researchers with extensive experience in survey research.

The RLMS is the first nationally representative random sample for Russia, albeit a highly clustered one. Data have been collected eight times since 1992. During the first phase of the project, in 1992–93, the RLMS collected four rounds of data. This periodicity represents a compromise between the interests of the major parties and partially relates to the availability of funds from both the Russian and US sides. In the second phase of the project, which began in 1994 and is ongoing, the RLMS has collected four more rounds of data. Of the 7,200 targeted households, 6,334 provided data for Round I (17,154 individuals, of which 4,148 are aged 55 and older). This is a response rate of 88.8 per cent. An additional 40 households (or less than 1 per cent of the sample) refused to participate in Round II interviews, while a number of Round I refusals agreed to be surveyed for Round II. In Round II, the target sample size was set at 4,000: however, the number of households drawn into the sample was inflated to 4,718 to allow for a non-response rate of approximately 15 per cent. The new RLMS sample was smaller, but the number of primary sampling units was doubled to enhance the representativeness of the survey. A variety of approaches have been used to reduce subsequent loss to follow-up (including *honoraria* to respondents and training interviewers to be courteous and respectful).

RLMS datasets at the individual and household levels are presently being made available via the World Wide Web: http://www.cpc.unc.edu/projects/rlms/data/rlmsform.html

The user has to complete a form and then check the datasets which she/he would like to receive. They will then receive an email with instructions for retrieving the dataset(s) via FTP.

To safeguard the confidentiality of RLMS respondents, such datasets do not include community-level data, which might be used in an attempt

to deduce location. Community-level data can, however, be useful to legitimate researchers studying regional differences in outcomes. In order for RLMS project staff to make them available in a manner that meets the requirements for the ethical treatment of human subjects set forth by the Institutional Review Board at the University of North Carolina at Chapel Hill, potential users of community-level data must agree to some guidelines and restrictions. By signing this agreement, the person requesting community-level data agrees to abide by all listed guidelines and restrictions, and acknowledges that any violation of the terms of this agreement may result in punitive legal action. Within approximately two weeks of receiving a completed copy of this agreement, RLMS staff will notify the researcher of their decision whether or not to approve the nature of the proposed research and the means by which the researcher will restrict access to confidential data. If these items are approved, within another two weeks, RLMS staff will send the requested data on diskette to the researcher, in SAS XPORT format, via US mail.
Web sites: http://www.cpc.unc.edu/projects/rlms/project/study.html
http://www.cpc.unc.edu/projects/rlms/project/sampling.html
http://www.cpc.unc.edu/projects/rlms/project/scheduling.html

The Swiss Household Panel

The Swiss Household Panel 'Vivre en Suisse – Leben in der Schweiz' (SHP) is an annual longitudinal panel survey financed by the Swiss National Science Foundation, the University of Neuchâtel and the Swiss Federal Statistical Office (SFSO). The SHP survey is a multi-purpose longitudinal study, set up to observe (gross) social change at individual and household level and the validation of causal hypotheses (using the temporal succession of events). Data are gathered at both the houshold level (characteristics of household members, household size, type of accommodation, etc.) and at the level of the persons living in the household (such as education, employment status, opinions).

A representative sample of the Swiss population (5,074 households) was recruited and interviewed in autumn 1999 (1st wave), resulting in the collection of individual data from 7,799 persons aged 14 years and older. At the household level the net response rate is 61 per cent and may be considered a good response rate for panel studies. All members of these households are to be re-interviewed annually for the next four years. An extension of the survey is planned. All interviews are made in German, French and Italian by means of the CATI technique. The data are readily available to all researchers upon signing a contractual agreement. Grants are offered for data analysis with SHP.
Web site: http://www.unine.ch/sm/

The Panel Comparability Project

The Panel Comparability Project (PACO) is a centralised approach, designed to create a set of comparable variables across a number of domains and countries to facilitate cross-national longitudinal research.

The PACO consists of:

1 the PACO archive;
2 the PACO database.

The PACO first set up a data archive of existing household panels in Europe and in the USA. Currently the PACO Panel Archive includes original panel datasets from 10 countries, as shown in Table A2.13. The PACO Archive contains original (not harmonised) variables, but the original data have been transformed from different platforms and formats into one common format: SPSS system files for Windows on the PC. The process of making data comparable is carried out by creating harmonised and consistent variables and files.

The PACO database contains comparable variables transformed according to a common plan and was built up by using standardised international classifications where available. Thus, it increases the accessibility and use of panel data for research and facilitates comparative cross-national and longitudinal research both on processes and on the dynamics of policy issues such as labour force participation, income distribution, poverty, social exclusion, problems of the elderly, etc. Each country file is sufficiently anonymised and can therefore be rated as a public use file. All files are held in a relational database structure. The data are stored as system files for the statistical package SPSS for Windows, containing identical variable names, labels, values and data structures. The complete database is 250 MB and is available on a CD-ROM. The PACO Database can be linked to a collection

Table A2.13 Available countries in the PACO Data Archive

Country	*Available years*	*Source*
Belgium	1992	PSBH
France (Lorraine)	1985–90	ESEML
Germany	1984–97	GSOEP
Hungary	1992–97	HHP
Luxembourg	1985–94	PSELL
Poland	1987–90; 1994–96	PHP
Spain (Galicia)	1992–93	GES
Sweden	1984, 1986, 1988, 1991	HUS
USA	1968–92	PSID
UK	1991–98	BHPS

Source: http://www.ceps.lu/paco/pacopres.htm

Table A2.14 Available countries and years in the PACO Database

Country	*Reference year*	*No. of households/persons*
France/Lorraine	1985–90	2,100/7,500
Germany	1984–96	5,900/12,200
Hungary	1992–94	2,100/5,800
Luxembourg	1985–94	2,000/6,000
Poland	1987–90; 1994–96	3,700/12,600
Spain/Galicia	1992–93	1,800/6,500
USA	1983–87; 1992–93	6,800/19,400
UK	1991–97	5,500/13,800

Source: http://www.ceps.lu/paco/pacopres.htm

of macro data. A set of macro variables have been extracted from the Eurostat CD for the year 1993 and from other statistical sources. The macro data are accessible from SPSS and can be matched with the PACO files. The relevant parts of the Mutual Information System on Social Security (MISSOC) publications about social security have been compiled and integrated into the PACO documentation system. The information available makes it possible to link original variables from national panel studies with the MISSOC data; on the other hand, it is also possible to retrieve the MISSOC information about selected PACO variables.

National Documentation about the original panel studies has been collected at the PACO data centre. Parts of this documentation are available in paper form (questionnaires, handbooks), other parts are available as WordPerfect files and as meta-data programs on the PC. Researchers using PACO have to sign an agreement concerning data use and privacy regulations. They are obliged to submit their research papers containing PACO results for inclusion into the working paper series of CEPS/INSTEAD. Guest researchers – while they are working at CEPS/INSTEAD – can use all data (both PACO Data Archive and PACO Database) and documentation available on site. At present the distribution of PACO data to outside users is restricted. The PACO Database (containing harmonised data and documentation) can be accessed by outside users, but not the PACO Panel Archive (containing the original data).
Web site: http://www.ceps.lu/paco/pacopres.htm

The PSID-GSOEP Equivalent Data File

The PSID-GSOEP Equivalent Data File is the result of a combined effort by the German Institute for Economic Research (DIW); the Centre for Policy Research at Syracuse University and the University of Michigan to provide equivalent variables suitable for cross-national analysis. It involves the PSID

(United States) and the GSOEP (Germany). The Equivalent Data File was developed because, although both these surveys gather similar data, the PSID and the GSOEP use different methods to collect their information and, consequently, it is difficult to compare the original two files directly. The PSID-GSOEP database is, thus, the product of an attempt to render two sets of data more homogeneous: the first version covers the years 1984–89; the second, with 11 waves, continues the work of standardisation up to 1994. PSID data on more than 25,000 individuals and 7,000 households and GSOEP data on over 17,000 individuals and 5,000 households are included.

Two identically formatted rectangular files are provided, one for the PSID and the other for the GSOEP. The available variables can be grouped under the following headings: demography (8); labour force (3); income (8); macro-economic indicators (1); weighting (5); organisational variables (3). Variables have been made comparable and the descriptions of the algorithms used to create this comparability are provided in the codebooks associated with the data files.

The data are designed to allow cross-national researchers, with little experience in panel data analysis, to access a simplified version of these panels, while also providing experienced panel data users with guidelines for formulating equivalent variables across countries. Most importantly, the equivalent data file provides a set of constructed variables that are not directly available on either of the two surveys and that are combinations of variables found in the original PSID and GSOEP datasets. Since the Equivalent Data File can be merged with the original surveys, PSID-GSOEP users can also easily incorporate these constructed variables into current analyses (Daly, 1994).

Web site: http://dpls.dacc.wisc.edu/apdu/gsoep_cd_data.html

The European Panel Analysis Group

The European Panel Analysis Group (EPAG) is a consortium of European social and economic researchers who have, since 1990, been collaborating in the development and analysis of HPSs in the EU. Most recently, EPAG has been engaged in the study of flexible labour and its impact on earnings and poverty, under a Eurostat contract, and on a programme of research on social exclusion, as part of the EU's Targeted Socio-Economic Research programme. The group has set up new comparative datasets based on five-year sequences of the British, German and Dutch national household panels and is, also, analysing early data from the ECHP. Most of the research to date has been in the fields of family formation, employment, household income and 'deprivation'. The group has recently been awarded a grant under the EU's Fifth Framework Programme, 'Improving Human Potential

and the Socio-Economic Knowledge Base', to undertake studies of the processes of change in the domains of family structure, employment, household income and living standards. This project – 'The Dynamics of Social Change in Europe' – began in March 2000, and is based primarily on the quantitative analysis of ECHP data.

The EPAG dataset can be accessed through the ECASS programme – which is a Large Scale Facility for the Social Sciences that offers access to files held in the Data Archive of the University of Essex.
Web site: http://www.irc.essex.ac.uk/epag/index.php

The Consortium of Household Panels for European Socio-Economic Research

The aim of the Consortium of Household Panels for European Socio-Economic Research (CHER) is to create an international comparative micro database containing longitudinal datasets from many European national household panels and from the country datasets available in the ECHP. The database will be supplemented with data from the US and from Canada. All this will be complemented by key information from existing macro/institutional datasets linked to the comparative database and supported by utilities for panel analyses. The final CHER database will contain harmonised and consistent variables and identical data structures for each country included. The co-ordinator of the Consortium is the Centre d'Études de Populations, de Pauvreté et de Politiques Socioeconomique in Luxembourg (CEPS/INSTEAD). The project partners are: Belgium, France, Germany, Greece, Hungary, Italy, Luxembourg, The Netherlands, Poland, Spain and Switzerland.
Web site: http://www.kub.nl/~fsw_2/asz/tisser/research/Cher.htm

Bibliography

Abbott, A. (1995) 'Sequence Analysis: New Methods for Old Ideas', *Annual Review of Sociology* 21, 93–113.

Abbott, A. (1997) *Program Optimize*. Chicago: University of Chicago. http://www.spc.uchicago.edu/users/abbot/optfdoc.htm.

Abbott, A. and Forrest, J. (1986) 'Optimal Matching Methods for Historical Sequences', *The Journal of Interdisciplinary History* 16, 471–94.

Abrams, P. (1982) *Historical Sociology*. West Compton House: Open Books Publishing Ltd.

Adler, M.C. (1992) 'The Future of SIPP for Analyzing Disability and Health', *Journal of Economic and Social Measurement* 18, 91–124.

Aitkin, M., Anderson, D. and Hinde, J. (1981) 'Statistical Modelling of Data on Teaching Styles (with discussion)', *Journal of The Royal Statistical Society* 144, series A, 148–61.

Allison, P.D. (1984) *Event History Analysis. Regression for Longitudinal Event Data*. London: Sage Publications.

Allison, P.D. (1995) *Survival Analysis Using SAS System: a Practical Guide*. Cary: Sas Institute.

Alwin, D.F. (1988) 'Structural Equation Models in Human Development and Aging', in K.W. Schaie, R.T. Campbell, W. Meredith and S.C. Rawlings (eds) *Methodological Issues in Aging Research*. New York: Springer Verlag, 70–170.

Anderson, N. (1923) *The Hobo: The Sociology of the Homeless Man*. Chicago: University of Chicago Press.

Arminger, G. and Mueller, F. (1990) *Lineare Modelle zur Analyse von Paneldaten*. Opladen: Westdeutscher Verlag.

Ashenfelter, O. and Solon, G. (1982) 'Longitudinal Labor Market Data. Sources, Uses and Limitations', in *Proceedings of the Conference on Labor Force Measurements for Policy Formulation*. Washington: US Government Printing Office, 109–26.

Ashworth, K., Hill, M.S. and Walker, R. (1992a) 'Patterns of Childhood Poverty: New Challenges for Policy', Centre for Research in Social Policy. Loughborough: Loughborough University of Technology. Working Paper 169.

Ashworth, K., Hill, M.S. and Walker, R. (1992b) 'Economic Disadvantages During Childhood', Centre for Research in Social Policy, Loughborough: Loughborough University of Technology. Working Paper 169.

Ashworth, K., Hill, M.S. and Walker, R. (2000) 'A New Approach to Poverty Dynamics', in D. Rose (ed.) *Researching Social and Economic Change. The Uses of Household Panel Studies*, Social Research Today Series. London: Routledge, 210–29.

Bailar, B. (1984) 'The Quality of Survey Data', in *Proceedings of The Survey Research Section*, American Statistical Association, 45–53.

Bailar, B. (1989) 'Information Needs, Surveys and Measurement Errors', in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds) *Panel Surveys*. New York: John Wiley, 1–24.

Bailey, K.D. (1994) *Methods of Social Research*. New York: The Free Press.

Bane, M.J. (1986) 'Household Composition and Poverty', in S.H. Danziger, D.H. Weinberg (eds) *Fighting Poverty: What Works and What Doesn't*. Cambridge, MA: Harvard University Press, 209–31.

Bane, M.J. and Ellwood, D.T. (1986) 'Slipping into and out of Poverty. The Dynamics of Spells', *The Journal of Human Resources* 21, 1, 1–23.

Barlow, R.E. and Proschan F. (1975) *Statistical Theory of Reliability and Life Testing*. New York: Holt Rinehart.

Bauman, Z. (1992) *Intimations of Postmodernity*. London: Routledge.

Beck, U. (1992) *Risk Society: Towards a New Modernity*. London: Sage.

Bell, D. (1973) *The Coming of Post-Industrial Society: A Venture in Social Forecasting*. London: Heinemann.

Bennett, N. (1976) *Teaching Styles and Pupil Progress*. London: Open Books.

Benus, J. (1975) 'Response Rates and Data Quality', in G.J. Duncan and J.N. Morgan (eds) *Five Thousand American Families: Patterns of Economic Progress*, vol. III. Ann Arbor, MI: Institute for Social Research.

Berrington, A. (1995) 'Measurement Errors in Retrospectives Union Histories: Implications for the Analysis of Partnership Formation', Paper presented at the European Population Conference, Milan.

Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A. and Sudman, S. (eds) (1991) *Measurement Errors in Surveys*. New York: John Wiley.

Bijleveld, C.C.J.H., Mooijaart, A., van der Kamp, L.J.Th. and van der Kloot, W.A. (1998) 'Structural Equation Models for Longitudinal Data', in C.C.J.H. Bijleveld *et al.* (eds) *Longitudinal Data Analysis. Designs, Models and Methods*. London: Sage Publications, 207–68.

Billari, F. (1998) 'Appunti di demografia sociale', University of Padua.

Billari, F. (1999) 'Sequence Analysis in Demography: Chances and Wishes', Paper presented at the Workshop on Longitudinal Research in Social Science – A Canadian Focus, Windereme Manor, London, Ontario, Canada, 25–27 October 1999.

Billari, F.C. and Piccarreta, R. (1999) 'Classifying and studying life-courses as sequences via monothetic divisive algorithms: an experiment', *Book of Short Papers CLADAG 99*, Classification and Data Analysis Group, Italian Statistical Society, Rome, 253–6.

Billari, F. and Rosina, A. (1999) 'Riflessioni sull'utilizzo dell'Event History Analysis in demografia: dalla teoria agli strumenti, dagli strumenti alla teoria', Paper presented at 'Giornate di Studio sulla Popolazione', Florence, 7–9 January.

Biorcio, R. and Pagani, S. (1997) *Introduzione alla ricerca sociale*. Rome: La Nuova Italia Scientifica.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975) *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

Blossfeld, H.P. (ed.) (1995) *The New Role Of Women. Family Formation in Modern Societies*. Boulder, CO: Westview Press.

Blossfeld, H.P. and Rohwer, G. (1995) *Techniques of Event History Modeling. New Approaches to Causal Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Blossfeld, H.P. and Rohwer, G. (1997) 'Causal Inference, Time and Observation Plans in the Social Sciences', *Quality and Quantity* 31, 361–84.

Blossfeld, H.P., Hamerle, A. and Mayer, K.U. (1989) *Event History Analysis. Statistical Theory and Application in the Social Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bound, J., Duncan, G.J., Laren, D. and Oleinick, L. (1991) 'Poverty Dynamics in Widowhood', *Journal of Gerontology* 46, 3, 115–24.

Bower, G., Black, J. and Turner, T. (1979) 'Scripts in Comprehension and Memory', *Cognitive Psychology* 11, 177–220.

Bradburn, N.M., Rips, L.J. and Shevell, S.K. (1987) 'Answering Autobiographical Questions: The Impact of Memory and Inferences on Survey', *Science* 236, 157–61.

Brandolini, A. and Cannari, L. (1994) 'Methodological Appendix: The Bank of Italy's Survey of Household Income and Wealth', in A. Ando, L. Guiso and I. Visco (eds) *Saving and the Accumulation of Wealth. Essays on Italian Household and Government Saving Behavior*. Cambridge: Cambridge University Press, 369–86.

Brannen, J. (ed.) (1992) *Mixing Methods: Qualitative and Quantitative Research*. Aldershot: Avebury.

Brückner, E. (1995) *Surveys don't Lie, People do? An Analysis of Data Quality in a Retrospective Life-course Study*, Materialen aus der Bildungsforschung n. 50, Berlin: Max-Planck-Institut für Bildungsforschung.

Brückner, E. and Mayer, K.U. (1997) 'Collecting Life History Data. Experiences from the German Life History Study', in J.Z. Giele and G.H. Elder Jr (eds) *Methods of Life-course Research: Qualitative and Quantitative Approaches*. London: Sage Publications, 152–81.

Bryk, A.S. and Raudenbush, S.W. (1992) *Hierarchical Linear Models*. Newbury Park: Sage Publications.

Bryman, A. (1988) *Quantity and Quality in Social Research*. London: Routledge.

Buck, N., Gershuny, J., Rose, D. and Scott, J. (eds) (1994) *Changing Households: the BHPS 1990 to 1992*. ESRC Research Centre on Micro-Social Change. Colchester: University of Essex.

Bulmer, M. (1983) *Sociological Research Methods: An Introduction*, 2nd edn. London: Macmillan, 8–12.

Burgess, R.D. (1989) 'Major Issues and Implications of Tracing Survey Respondents' in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds) *Panel Surveys*. New York: John Wiley, 52–74.

Burkhauser, R.V. (1991) 'An Introduction to The German Socio-Economic Panel for English Speaking Researchers'. Syracuse, NY: Syracuse University. Project Paper 1.

Burkhauser, R.V. and Duncan, G.J. (1988) 'Life Events, Public Policy, and the Economic Vulnerability of Children and the Elderly', in J.L. Palmer, T.M. Smeeding and B.B. Torrey (eds) *The Vulnerable*. Washington, DC: The Urban Institute Press, Ch. 4.

Burkhauser, R.V. and Duncan, G.J. (1991) 'United States Public Policy and The Elderly. The Disproportionate Risk to the Well-Being of Women', *Journal of Population Economics* 4, 217–31.

Burkhauser, R.V. and Wagner, G.G. (eds) (1990) *The German Socio-Economic Panel (GSOEP). User Handbook*, Version 4. Berlin: DIW.

Burkhauser, R.V., Butrica, B.A. and Daly, M.C. (1995) 'The Syracuse University Panel Study of Income Dynamics and German Socio-Economic Panel Equivalent Data File', Cross-National Studies in Aging, Center for Policy Research. The Maxwell School. Syracuse, NY: Syracuse University. Program Project Paper 25.

Burkhauser, R.V., Butrica, B.A. and Daly, M.C. (1999) 'The PSID-GSOEP Equivalent File: A Product of Cross-National Research', in W. Voges (ed.) *Dynamic Approaches to Comparative Social Research: Recent Developments and Applications*. Aldershot: Ashgate Publishing Ltd, 53–66.

Butrica, B.A. (1996a) 'Getting Started with The Public-Use Version of The German Socio-Economic Panel (GSOEP)', The National Institute On Aging, The Well-Being

Of The Elderly in A Comparative Context. Syracuse, NY: Syracuse University. Program Project Paper #1–P01–Ag09743–01.

Butrica, B.A. (1996b) 'Imputation Methods For Filling in Missing Values in the PSID-GSOEP Equivalent File 1980–1994', The National Institute On Aging, The Well-Being Of The Elderly in A Comparative Context. Syracuse, NY: Syracuse University. Program Project Paper #1–P01–Ag09743–01.

Bynner, J. (1996) *Use of Longitudinal Data in the Study of Social Exclusion. A Report for the Organisation for Economic Co-operation and Development*, Social Statistics Research Unit. London: City University.

Callan, T., Nolan, B., Whelan, B.J., Whelan, C.T. and Williams, J. (1996) *Poverty in the 1990s. Evidence from the 1994 Living in Ireland Survey*. Dublin: The Economic and Social Research Institute and Combat Poverty Agency, 31–53.

Campbell, D.T. and Stanley, J.C. (1963) *Experimental and Quasi-experimental Designs for Research*. Chicago: Rand McNally.

Campbell, R.T. (1996) *Analysis of Panel Data*. Chicago: University of Illinois.

Cantillon, B. (1990) *Mutations sociodemographiques et securité sociale*, 'Bulletin de documentation', Belgian Ministry of Finance, November–December, 228–57.

Cantor, N. and Mischel, W. (1977) 'Traits as Prototypes: Effects on Recognition Memory', *Journal of Personality and Social Psychology* 35, 38–48.

Capursi, V. (1993) 'Some Issues in Longitudinal Data Analysis with Categorical Variables'. University of Palermo: Istituto di Statistica Sociale e Scienze Demografiche e Biometriche. Paper.

Carroll, G.C. (1983) 'Dynamic Analysis of Discrete Dependent Variables: A Didactic Essay', *Quality and Quantity* 17, 425–60.

Center for Human Resource Research (1981) *The National Longitudinal Surveys Handbook*. Columbus: Ohio State University.

Centre for Longitudinal Studies (1999), *Research Review 1999*, Institute of Education, University of London, London.

Citro, C. and Kalton, G. (eds) (1993) *The Future of the Survey of Income and Program Participation*. Washington: National Academy Press.

Coe, R.D. (1988) 'A Longitudinal Examination of Poverty in the Elderly Years', *The Gerontologist* 28, 4, 540–4.

Coe, R.D., Duncan, G.J. and Hill, M.S. (1982) 'Dynamic Aspects of Poverty and Welfare Use in the United States', Paper presented at the Conference on Problems on Poverty, Clark University, August.

Coleman, J.S. (1981) *Longitudinal Data Analysis*. New York: Basic Books.

Comte, A. (1842) *Cours de philosophie positive*, vol. I. Paris: A. Costes.

Copeland, A.P. and White, K.M. (1991) *Studying Families*, Applied Social Research Methods Series, vol. 27. London: Sage Publications.

Corbetta, P. (1999) *Metodologia e tecniche della ricerca sociale*. Bologna: Il Mulino.

Corder, L.S. and Horvitz, D.G. (1989) 'Panel Effects in the National Medical Care Utilization and Expenditure Survey', in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds) *Panel Surveys*. New York: John Wiley, 304–18.

Cotton, C. and Giles, P. (1998) 'The Seam Effect in the Survey of Labour and Income Dynamics', The Income and Labour Dynamics Working Paper Series, Catalogue 98–18, November. Ottawa: Statistics Canada.

Courgeau, D. and Lelièvre, E. (1988) *Analyse démographique des biographies*. Paris: INED.

Cox, R.D. (1972) 'Regression Models and Life Tables', *Journal of The Royal Statistical Society* 34, series B, 187–220.

Cox, R.D. (1975) 'Partial Likelihood', *Biometrika* 62, 269–76.

Dale, A. and Davies, R.B. (1994a) *Analyzing Social and Political Change. A Casebook for Methods*. London: Sage Publications.

Dale, A. and Davies, R.B. (1994b) 'Modelling Trends in Multi-way Tables, Editor's Introduction', in A. Dale and R.B. Davies (eds) *Analyzing Social and Political Change. A Casebook for Methods*. London: Sage Publications, Ch. 3, 41–2.

Daly, M.C. (1994) 'The Syracuse University Panel Study of Income Dynamics and German Socio-Economic Panel Equivalent Data File', Cross-National Studies on Aging. Syracuse, NY: Syracuse University. Program Project Paper 16.

Danmarks Statistik (1991) *IDA – en integreret database for arbejdsmarkedsforskning*, Hovedrapport.

Danmarks Statistik (1993) *Registerbaseret arbejdsstyrkestatistik, ultimo november 1991*, Arbejdsmarked.

Danmarks Statistik (1994) *Personstatistik i Danmark. Et registerbaseret statistiksystem*. Copenhagen: Danmarks Statistiks Trykkeri.

David, M. (1991) 'The Science of Data Sharing: Documentation', in J.E. Sieber (ed.) *Sharing Social Science Data: Advantages and Challenges*. London: Sage Publications, 91–115.

Davies, R.B. (1987) 'The Limitations of Cross-Sectional Analysis', in R. Crouchley (ed.) *Longitudinal Data Analysis*. Aldershot: Avebury.

Davies, R.B. (1994) 'From Cross-Sectional to Longitudinal Analysis', in A. Dale and R.B. Davies (eds) *Analysing Social and Political Change. A Casebook of Methods*. London: Sage Publications, 20–40.

Davies, R.B. and Dale, A. (1994) 'Introduction', in A. Dale and R.B. Davies (eds) *Analysing Social and Political Change. A Casebook of Methods*. London: Sage Publications.

D'Alessio, G. and Faiella, I. (eds) (2000) *I bilanci delle famiglie italiane nell'anno 2000*, Supplementi al Bollettino Statistico, Note metodologiche e informazioni statistiche, anno XII, n. 6. Rome: Banca d'Italia.

De Graaf, N.D. (1999) 'Event History Data and making a History out of a Cross-sectional Data', in E. Ruspini (ed.) *Longitudinal Analysis: A Bridge between Quantitative and Qualitative Social Research*, Special Issue of *Quality and Quantity* August, 261–76.

Deleeck, H., Cantillon, B., Meulemans, B. and Van den Bosch, K. (1992) 'Some longitudinal results of the Belgian Socio-Economic Panel', *Journal of Income Distribution* 2, 2, 57–73.

Delhausse, B. (1992) *La Pauvrete en Belgique: Utilisation de donnees de panel*. Université de Liege.

Dex, S. (1991) 'The Reliability of Recall Data: A Literature Review', Occasional Papers of the ESRC Research Centre on Micro-Social Change. Colchester: University of Essex. Paper 6.

Dippo, C.S. (1989) 'The Use of Cognitive Laboratory Techniques for Investigating Memory Retrieval Errors in Retrospective Surveys', *Bulletin of the International Statistical Institute*, vol. LIII, Book 2, 363–82.

Ditch, J., Barnes, H., Bradshaw, J. and Kilkey, M. (1998) *Synthesis of National Family Policies*, EU Observatory on National Family Policies, University of York.

Dixon, W.J. (1988) *BMDP Statistical Software Manual*, vol. 2. Berkeley, CA: University of California Press.

Doolittle, R. (1990) *Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences*, vol. 183 of Methods in Enzymology. New York: Academic Press.

Draper, T.W. and Marcos, A.C. (eds) (1990) *Family Variables. Conceptualization, Measurement, and Use*. London: Sage Publications.

Duncan, G.J. (1989) *Panel Studies of Poverty: Prospects and Problems*, Survey Research Center, Ann Arbor: The University of Michigan.

Duncan, G.J. (1992) 'Household Panel Studies: Prospects and Problems', Working Papers of the European Scientific Network on Household Panel Studies. Colchester: University of Essex. Paper 54.

Duncan, G.J. (2000) 'Using Panel Studies to Understand Household Behaviour and Well-Being', in D. Rose (ed.) *Researching Social and Economic Change. The Uses of Household Panel Studies*, Social Research Today Series. London: Routledge, 54–75.

Duncan, G.J. and Kalton, G. (1987) 'Issues of Design and Analysis of Surveys Across Time', *International Statistical Review* 55, 97–117.

Duncan, G.J., Juster, F.T. and Morgan, J.N. (1987) 'The Role of Panel Studies in Research on Economic Behavior', *Transportation Research* 21A, 249–63.

Duncan G.J., Gustafsson, B., Hauser, R., Schmaus, G., Messinger, H., Muffels, R., Nolan, B. and Ray, J.C. (1993) 'Poverty Dynamics in Eight Countries', Working Papers of the European Scientific Network on Household Panel Studies. Colchester: University of Essex. Paper 5.

Edwards, W.S., Sperry, S. and Edwards, B. (1993) 'Using CAPI in a Longitudinal Survey. A Report from the Medicare Current Beneficiary Survey', in *Proceedings: Symposium 92, Design and Analysis of Longitudinal Surveys*. Ottawa: Statistica Canada, 21–30.

Egidi, V. and Costa, C. (1998) *Indagini e basi informative per lo studio di dati biografici*, comunicazione inviata alla XXXIX Riunione Scientifica della Società Italiana di Statistica, sessione specializzata *Analisi di biografie per processi demografico-sociali*, Sorrento.

Eisenhower, D., Mathiowetz, N.A. and Morganstein, D. (1991) 'Recall Error: Sources and Bias Reduction Techniques', in P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz and S. Sudman (eds) *Measurement Errors in Surveys*. New York: Wiley.

Elder, G.H. Jr. (1974) *Children of the Great Depression. Social Change in Life Experience*. Chicago: University of Chicago Press (1999, reissued as 25th Anniversary Edition, Boulder, CO: Westview Press).

Elder, G.H. Jr. (1985) 'Perspectives on the Life-course', in G.H. Jr. Elder (ed.) *Life-course Dynamics. Trajectories and Transitions, 1968–1980*. Ithaca and London: Cornell University Press, 23–49.

Elder, G.H. Jr. (1992) 'Life-course', in E.F. Borgatta and M.L. Borgatta (eds) *Encyclopedia of Sociology*, vol. 3. New York: Macmillan, 1120–30.

Elder, G.H. Jr. and Liker, J.K. (1982) 'Hard Times in Women's Lives: Historical Influences across 40 Years', *American Journal of Sociology* 88, 241–69.

Elder, G.H. Jr., Modell, J. and Parke, R.D. (eds) (1993) *Children in Time and Place: Developmental and Historical Insights*. Cambridge: Cambridge University Press.

Elster, J. (1989) *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

Engel, U. and Reinecke, J.R. (1994) *Panelanalyse*. New York: De Gruyter.

Engel, U. and Reinecke, J.R. (1996) 'Introduction', in U. Engel and J.R. Reinecke (eds) *Analysis of Change. Advanced Techniques in Panel Data Analysis*. Berlin and New York: Walter de Gruyter.

Erikson, R. and Åberg, R. (eds) (1987) *Welfare in Transition. A Survey of Living Conditions in Sweden 1968–1981*. Oxford: Oxford University Press.

Ermisch, J. (2000) 'Using Panel Data to Analyse Household and Family Dynamics', in D. Rose (ed.) *Researching Social and Economic Change. The Uses of Household Panel Studies*, Social Research Today Series. London: Routledge, 230–49.

Eurostat (1996a) *The European Community Household Panel (ECHP): Survey Methodology and Implementation*, vol. 1. Luxembourg: Statistical Office of the European Communities.

Eurostat (1996b) *The Household Panel Newsletter*, Theme 3, Population and Social Conditions, Series B, Short-term Statistics, Luxembourg.

Eurostat (1997) *Response Rate for the First Three Waves of the ECHP*, ECHP Working Goup, Doc PAN 92/97. Luxembourg: Statistical Office of the European Communities.

Eustis, N.N., Clark, R.F. and Adler, M.C. (1995) 'Disability Data for Disability Policy: Availability, Access and Analysis', Background Paper for the HHS/ASPE Office of Disability, Aging, and Long-Term Care, Policy's third 'Expert' meeting, revised version, August. http://aspe.hhs.gov/daltcp/home.htm.

Fay, R.E. (1989) 'Estimating Nonignorable Nonresponse in Longitudinal Surveys through Causal Modeling', in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds) *Panel Surveys*. New York: John Wiley, 375–99.

Featherman, D.L. (1980) 'Retrospective Longitudinal Research: Methodological Considerations', *Journal of Economics and Business* 32, 152–69.

Federal Committee on Statistical Methodology – Subcommittee on Federal Longitudinal Surveys (1985) *Federal Longitudinal Surveys*, Statistical Policy Working Paper 13. Washington, DC: U.S. Government Printing Office. http://www.fcsm.gov/working-papers/sw13.html.

Fergusson, D.M. and Horwood, L.J. (1988) 'Structural Equation Modelling of Measurement Processes in Longitudinal Data', in M. Rutter (ed.) *Studies of Psychosocial Risks: The Power of Longitudinal Data*. Cambridge: Cambridge University Press, 325–53.

Fideli, R. (1999) *La comparazione*, Collana Metodologia delle Scienze Umane, vol. 10. Milan: Franco Angeli.

Finkel, S.E. (1995) *Causal Analysis with Panel Data*, Sage University Paper Series on Quantitative Applications in the Social Sciences, vol. 105. Thousand Oaks: Sage Publications.

Firebaugh, G. (1997) *Analyzing Repeated Surveys*, Sage University Paper Series on Quantitative Applications in the Social Sciences (QASS), n. 115. London: Sage Publications.

Flood, L., Klevmarken, A. and Olovsson, P. (1997) *Household Market and Non Market Activities (HUS). Survey Description*, vol. III. Uppsala University: Department of Economics.

Francis, B., Green., M. and Payne, C. (1993) *The GLIM System Release 4 Manual*. Oxford: Oxford University Press.

Freed Taylor, M. (2000) 'Dissemination Issues for Panel Studies: Metadata and Documentation', in D. Rose (ed.) *Researching Social and Economic Change. The Uses of Household Panel Studies*, Social Research Today Series. London: Routledge, 146–62.

Freed Taylor, M., Brice, J. and Buck, N. (eds) (1995) *BHPS User Manual. Introduction, Technical Report and Appendices*, vol. A. Colchester: University of Essex.

Freedman, D.A. (1987) 'As Others See Us: A Case Study in Path Analysis', *Journal of Educational Statistics* 12, 101–28.

Freedman, D.A. (1991) 'Statistical Models and Shoe Leather (with discussion)', *Sociological Methodology* 21, 291–358.

Freedman, D., Thornton, A. and Camburn, D. (1980) 'Maintaining Response Rates in Longitudinal Studies', *Sociological Methods and Research* 9, 87–9.

Freedman, D., Thornton, A., Camburn, D., Alwin, D. and Young-DeMarco, L. (1988) 'The Life History Calendar: A Technique for Collecting Retrospective Data', *Sociological Methodology* 18, 37–68.

Frey, J.H. (1989) *Survey Research by Telephone*, Sage Library of Social Research, vol. 150. London: Sage Publications.

Frick, J. (1998) Handout prepared for GSOEP Summer Workshop 29–30 June 1998. Berlin, German Institute for Economic Research (DIW). Material distributed to GSOEP users on CD-ROM.

Fuchs, W. (1984) *Biographische Forschung: Eine Einführung in Praxis und Methoden*. Opladen: Westdeutscher Verlag.

Fuller, W.A. (1987) *Measurement Error Models*. New York: John Wiley.

Furu, M. (1995) 'Growth Conditions and Education in Relation to Health in a Life-long Perspective', in H. Daun, M. O'Dowd and S. Zhao (eds) *The Role of Education in Development. From Personal to International Arenas*. Stockholm University: Institute of International Education, Studies in Comparative and International Education.

Furu, M. and Hellström, Z. (1996) 'Work Situation and Health: A Prospective Ten-Year Study of Gender Differences in An Elderly Normal Population', Paper presented at the Fourth International Conference in Behavioral Medicine, Washington.

Gallie, D. and Paugam, S. (eds) (2000) *Welfare Regimes and the Experience of Unemployment in Europe*. Oxford: Oxford University Press.

Gallino, L. (1993) *Dizionario di Sociologia*. Milan: Tea.

Gershuny, J. (1998) 'Thinking Dynamically: Sociology and Narrative Data', in L. Leisering and R. Walker (eds) *The Dynamics of Modern Society*. Bristol: The Policy Press, 34–48.

Gershuny, J. (2000) 'Time Budgets, Life Histories and Social Position', in E. Ruspini (ed.) *Longitudinal Analysis: A Bridge between Quantitative and Qualitative Social Research*, Special Issue of *Quality and Quantity*, August, 277–89.

Gershuny, J. and Buck, N. (2000) *ESRC National Strategy for Longitudinal Studies: Enquiry into Social Scientists' Requirements for Longitudinal Data*, UK Longitudinal Studies Centre. Colchester: University of Essex. http://www.iser.essex.ac.uk/activities/NSLS-Enquiry.php.

Gershuny, J. and Buck, N. (2001) *Understanding Panel Data*. London: Sage.

Ghellini, G. (1994) 'Controllo e analisi della qualità dei dati in un'indagine panel sulle condizioni di vita', Paper presented at the Second International Conference 'Metodi Quantitativi per le Scienze Applicate', Certosa di Pontignano (Siena), 27–29 June.

Ghellini, G. and Trivellato, U. (1996) 'Indagini panel sul comportamento socio-economico di individui e famiglie: una selezionata rassegna di problemi ed esperienze', in C. Quintano (ed.) *Scritti di Statistica Economica 2*. Naples: Rocco Curto Editore.

Giddens, A. (1990) *The Consequences of Modernity*. Cambridge: Polity Press.

Giele, J.Z. and Elder, G.H. Jr. (eds) (1998) *Methods of Life-course Research. Qualitative and Quantitative Approaches*. Thousand Oaks: Sage Publications.

Gilbert, N. (1993) *Analyzing Tabular Data. Loglinear and Logistic Models for Social Researchers*. London: University College London Press.

Glenn, N.D. (1977) *Cohort Analysis*. Beverly Hills: Sage Publications.

Gobo, G. (1997) *Le risposte e il loro contesto. Processi cognitivi e comunicativi nelle interviste standardizzate*, vol. 8, Collana Metodologia delle Scienze Umane. Milan: Franco Angeli.

Goldthorpe, J. (in collaboration with Llewelyn, C. and Payne, C.) (1987) *Social Mobility and Class Structure in Modern Britain*, 2nd edn. Oxford: Oxford University Press.

Goodin, R.E., Headey, B., Muffels, R. and Dirven, H.-J. (1999) *The Real Worlds of Welfare Capitalism*. Cambridge: Cambridge University Press.

Goodman, L.A. (1973) 'Causal Analysis of Data from Panel Studies and Other Kinds of Surveys', *American Journal of Sociology* 78, 1135–91.

Gross, A.J. and Clark, V.A. (1975) *Survival Distributions: Reliability Applications in the Biomedical Sciences*. New York: John Wiley.

Groves, R.M. and Kahn, R.L. (1979) *Surveys by Telephone: A National Comparison with Personal Interviews*. New York: Academic Press.

Groves, R.M., Biemer, P.P., Lyberg, L.E., Massey, J.T., Nichols, W.L. and Walksberg, J. (1988) *Telephone Survey Methodology*, Wiley Series in Probability and Mathematical Statistics, Applied Probability and Statistics, 0271–6356. New York: John Wiley.

Gustafsson, B. (1997) 'The Swedish Income Panel (SWIP)', Application to the Swedish Council for Planning and Co-ordination of Research (FRN), Longitudinal Research, Paper, University of Gothenburg, March.

Haferkamp, H. and Smelser, N.J. (eds) (1992) *Social Change and Modernity*. Berkeley and Los Angeles: University of California Press.

Hagenaars, J.A. (1990) *Categorical Longitudinal Data. Log-Linear Panel, Trend and Cohort Analysis*. London: Sage Publications.

Hagenaars, J.A. and Luijkx, R. (1990) 'LCAG; Latent Class Models and Other Loglinear Models With Latent Variables, manual LCAG', Working Paper Series 17, revised edition. Tilburg University: Department of Sociology.

Hakim, C. (1987) *Research Design. Strategies and Choices in the Design of Social Research*. London: Allen and Unwin.

Hallinan, M.T. (1997) 'The Sociological Study of Social Change', *American Sociological Review* 62, 1–11.

Halpin, B. and Chan, T.W. (1998) 'Class Careers as Sequences: An Optimal Matching Analysis of Work-Life Histories', *European Sociological Review* 14, 111–30.

Handl, J. (1988) *Berufschancen und Heiratsmuster von Frauen: Empirische Untersuchungen zu Prozessen sozialer Mobilitaet*. Frankfurt am Main: Campus Verlag.

Hantrais, L. (1996) 'Comparative Research Methods', *Social Research Update*, Issue 13. Guildford: University of Surrey.

Hantrais, L. and Mangen, S. (1996) *Cross-National Research Methods in the Social Sciences*. London and New York: Pinter.

Hellström, Z. (1996) *The Malmö Study. Data Collection 1994*, Report 100. Stockholm University, Institute of International Education.

Herzog, A.R. and Rodgers, W.L. (1988) 'Interviewing Older Adults: Mode Comparison Using Data from a Face-to-face Survey and a Telephone Resurvey', *Public Opinion Quarterly* 52, 84–99.

Hill, M. and Jenkins, M.S. (1999) *Poverty Among British Children: Chronic or Transitory?*, Institute for Social and Economic Research. Colchester: University of Essex. Working Paper 99–23.

Hill, M.S. (1992) *The Panel Study of Income Dynamics. A User's Guide*, Guides to Major Social Sciences Data Bases. London: Sage Publications.

Hox, J.J. and Creft, I.G.G. (1994) 'Multilevel Analysis Methods', *Sociological Methods & Research* 22, 3, 283–99.

Hsiao, C. (1985) 'Benefits and limitations of panel data', *Econometric Reviews* 4: 121–88.

Hsiao, C. (1986) *Analysis of Panel Data*. Cambridge, MA: Cambridge University Press.

Huinink, J., Mayer, K.U., Diewald, M., Solga, H., Sorensen, A. and Trappe, H. (1995) *Kollektiv und Eigensinn: Lebensverlaufe in der DDR und danach*. Berlin: Akademie Verlag.

ISTAT (Istituto Nazionale di Statistica) (1993) 'Obiettivi, disegno e metodologia', vol. 1 in *Indagine Multiscopo sulle Famiglie*. Rome: ISTAT.

Jaccard, J. and Wan Choi, K. (1996) *LISREL approaches to interaction effects in multiple regression*. Thousand Oaks: Sage Publications.

Janson, C.-G. (1990) 'Retrospective Data, Undesirable Behavior, and the Longitudinal Perspective', in D. Magnusson and L.R. Bergman (eds) *Data Quality in Longitudinal Research*. Cambridge: Cambridge University Press.

Jarvis, S. and Jenkins, S. (1998) 'Marital Dissolution and Income Change: Evidence for Britain', in R. Ford and J. Millar (eds) *Private Lives and Public Responses: Lone Parenthood and Future Policy*. London: Policy Studies Institute.

Jarvis, S. and Jenkins, S. (2000) 'Low-income Dynamics in 1990s Britain', in D. Rose (ed.) *Researching Social and Economic Change. The Uses of Household Panel Studies*, Social Research Today Series. London: Routledge, 188–209.

Jedlowski, P. (1998) *Il mondo in questione. Introduzione alla storia del pensiero sociologico*. Rome: Carocci Editore.

Johansson, S. (1973) 'The Level of Living Survey: A Presentation', *Acta Sociologica* 16, 211–19.

Jones, K. (1993) 'Using Multilevel Models for Survey Analysis', *Journal of the Market Research Society* 38, 8, 249–65.

Jones, K. and Duncan, C. (1994) 'People and Places: the Multilevel Model as a General Framework for the Quantitative Analysis of Geographical Data', Paper prepared for the Workshop 'Spatial Modelling and GIS', Bristol, June.

Joshi, H. and Davies, H. (2002) 'Women's Incomes Over A Synthetic Lifetime', in E. Ruspini and A. Dale (eds) *The Gender Dimension of Social Change. The Contribution of Dynamic Research to the Study of Women's Life-courses*. Bristol: The Policy Press.

Jowell, R., Brooks, L., Prior, G. and Taylor, B. (1992) *British Social Attitudes: the 9th Report*. Aldershot: Avebury.

Kalbfleisch, J.D. and Prentice, R.L. (1980) *The Statistical Analysis of Failure Time Data*. New York: John Wiley.

Kalton, G. (1983) *Compensating for Missing Survey Data*. Ann Arbor: Institute for Social Research, University of Michigan.

Kalton, G. and Brick, M. (2000) 'Weighting in Household Panel Surveys', in D. Rose (ed.) *Researching Social and Economic Change. The Uses of Household Panel Studies*, Social Research Today Series. London: Routledge, 96–112.

Kalton, G. and Citro, C.F. (2000) 'Panel Surveys. Adding the Fourth Dimension', in D. Rose (ed.) *Researching Social and Economic Change. The Uses of Household Panel Studies*, Social Research Today Series. London: Routledge, 36–53.

Kalton, G. and Lepkowski, J. (1985) 'Following Rules in SIPP', *Journal of Social and Economic Measurement* 13, 319–29.

Kalton, G., Kasprzyk, D. and McMillen, D.B. (1989) 'Nonsampling Errors in Panel Survey', in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds) *Panel Surveys*. New York: John Wiley, 249–70.

Kasprzyk, D., Duncan, G.J., Kalton, G. and Singh, M.P. (eds) (1989) *Panel Surveys*. New York: John Wiley.

Kelly, J.R. and McGrath, J.E. (1988) *On Time and Method*. London: Sage Publications.

Kessler, R.C. and Greenberg, D.F. (1981) *Linear Panel Analysis*. London: Academic Press.

King, G. (1989) *Unifying Political Methodology. The Likelihood Theory of Statistical Inference*. Cambridge: Cambridge University Press.

Kish, L. (1986) 'Timing of Surveys for Public Policy', *Australian Journal of Statistics* 7, 1–12.

Kish, L. (1987) *Statistical Design for Research*. New York: John Wiley.

Klevmarken, A. and Olovsson, P. (1993) *Household Market and Non Market Activities (HUS). Procedures and Codes 1984–1991*. Stockholm: The Industrial Institute for Economic and Social Research.

Kline, R.B. (1998) *Principles and Practice of Structural Equation Modeling*. New York: Guilford Press.

Kohn, M.L. (ed.) (1989) *Cross-National Research in Sociology*. London: Sage Publications.

Laurie, H. (2000) 'From PAPI to CAPI: Consequences for Data Quality on the British Household Panel Survey', Paper presented at the Fifth International Conference on Logic and Methodology, University of Cologne, 3–6 October.

Laurie, H., Smith, R. and Scott, L. (1997) 'Strategies for Reducing Non-response in a Longitudinal Panel Survey', ESCR Research Centre on Micro-Social Change. Colchester: University of Essex. Working Paper.

Lavrakas, P.J. (1987) *Telephone Survey Methods. Sampling, Selection and Supervision*, Applied Social Research Methods Series, vol. 7. London: Sage Publications.

Lazarsfeld, P.F. (1940) 'Panel Studies', *Public Opinion Quarterly* 4, 122–8.

Lazarsfeld, P.F. (1948) 'The Use of Panels in Social Research', in *Proceedings of the American Philosophical Society* 42, 405–10.

Lazarsfeld, P.F. (1972) 'Mutual Relations Over Time of Two Attributes: A Review and Integration of Various Approaches', in M. Hammer, K. Salzinger and S. Sutton (eds) *Psychopathology*. New York: Basic Books, 461–80.

Lazarsfeld, P.F. and Fiske, M. (1938) 'The "Panel" as a New Tool for Measuring Public Opinion', *Public Opinion Quarterly* 2, 596–612.

Leisering, L. and Walker, R. (1998a) 'Preface', in L. Leisering and R. Walker (eds) *The Dynamics of Modern Society*. Bristol: The Policy Press, x–xvii.

Leisering, L. and Walker, R. (1998b) 'Making the Future: From Dynamics to Policy Agendas', in L. Leisering and R. Walker (eds) *The Dynamics of Modern Society*. Bristol: The Policy Press, 265–85.

Lemmens, R.M.M. (1991) *The Socio-Economic Panel Survey. Content, Design and Organisation*. Voorburg/Heerlen: Central Bureau of Statistics of the Netherlands, Statistics of Income and Consumption Branch.

Lepkowski, J. (1989) 'Treatment of Wave Nonresponse in Panel Surveys', in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds) *Panel Surveys*. New York: John Wiley, 348–74.

Lessler, J.T. and Kalsbeek, W. (1992) *Nonsampling Errors in Surveys*. New York: John Wiley.

Leth-Sørensen, S. (1997) *The IDA Database. A Longitudinal Database of Establishments and Their Employees*, The IDA Project, 3rd Division, Copenhagen: Danmarks Statistik.

Lillard, L. and Waite, L.J. (1995) 'Till Death Do Us Part: Marital Disruption and Mortality', *American Journal of Sociology* 100, 5, 1131–56.

Linton, M. (1982) 'Transformations of Memory in Everyday Life', in U. Neisser (ed.) *Memory Observed. Remembering in Natural Contexts*. San Fransisco: W. H. Freeman.

Little, R.J.A. (1988) 'Missing Data Adjustments in Large Surveys', *Journal of Business and Economic Statistics* 6, 287–96.

Little, R.J.A. and Su, H.-L. (1989) 'Item Non-Response in Panel Surveys', in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds) *Panel Surveys*. New York: John Wiley, 400–25.

Loftus, E.F. and Marburger, W. (1983) 'Since the Eruption of Mt. St. Helens, Has Anyone Beaten you up? Improving the Accuracy of Retrospective Reports with Landmark Events', *Memory and Cognition* 11, 114–20.

Loftus, E.F. and Palmer, J.C. (1974) 'Reconstruction of Automobile Destruction: An Example of the Interaction between Language and Memory', *Journal of Verbal Learning and Behavior* 13, 535–89.

Longford, N.T. (1993) *Random Coefficient Models*. Oxford: Clarendon Press.

MacCallum, R.C., Kim, C., Malarkey, W.B. and Kiecolt-Glaser, J.K. (1996) 'Studying Multivariate Change Using Multilevel Models and Latent Curve Models', *Multivariate Behavioral Research* 32, 2, 215–53.

Magnusson, D. and Bergman, L.R. (eds) (1990) *Data Quality in Longitudinal Research*. Cambridge: Cambridge University Press.

Magnusson, D., Bergman, L.R., Rudinger, G. and Torestad, B. (eds) (1991) *Problems and Methods in Longitudinal Research: Stability and Change*. Cambridge: Cambridge University Press.

Mandler, J.M. and Johnson, N.J. (1977) 'Remembrance of Things Passed: Story Structure and Recall', *Cognitive Psychology* 9, 111–51.

Mannheim, K. (1952) 'The Problem of Generations', in P. Kecskemeti (ed.) *Essays on the Sociology of Knowledge*. New York: Oxford University Press: 276–322.

Markus, G.B. (1979) *Analyzing Panel Data*, Sage University Paper Series on Quantitative Applications in the Social Sciences (QASS), 07–018. London: Sage Publications.

Marlier, E. (1999) *The EC Household Panel Newsletter*, Statistics in Focus, Population and Social Conditions, Theme 3, n. 16. Luxembourg: Eurostat.

Marradi, A. (1982) 'Forma e scopi della comparazione', in N. Smelser (ed.) *La comparazione nelle scienze sociali*. Bologna: Il Mulino.

Marradi, A. (1985) 'Natura, forma e scopi della comparazione', in D. Fisichella (ed.) *Metodo scientifico e ricerca politica*. Rome: Nuova Italia Scientifica, 293–321.

Marradi, A. (1990) 'Fedeltà di un dato, affidabilità di una definizione operativa', *Rassegna Italiana di Sociologia*, anno 31, 1, 55–96.

Martin, J. and Roberts, C. (1984b) *Women and Employment: Technical Report*. London: OPCS.

Martin, J. and Roberts, C. (1984a) *Women and Employment: a Lifetime Perspective*. London: HMSO.

Martini, A. (1989) 'Seam Effect, Recall Bias, and the Estimation of Labour Force Transition Rates from SIPP', in *Proceedings of the Section of Survey Research Methods*, American Statistical Association, 387–92.

Mastrovita, S. (1998) 'La Event History Analysis: i costrutti fondamentali, i problemi, le prospettive di applicazione', *Sociologia e Ricerca Sociale* 57, 74–103.

Mátyás, L. and Sevestre, P. (eds) (1996) *The Econometrics of Panel Data: Handbook of Theory and Applications*, 2nd revised edn. Dordrecht: Kluwer Academic Publishers.

Mayer, K.U. (1977) *Fluktuation und Umschichtung. Empirische Untersuchungen zu Strukturen sozialer Ungleichheit und Prozessen sozialer Mobilität in der Bundesrepublik Deutschland* (Habilitationsschrift). Mannheim: University of Mannheim.

Mayer, K.U. (1990) 'Lebensverlaufe und sozialer Wandel. Anmerkungen zu einem Forschungsprogramm', *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, Sonderheft 31. Opladen: Westdeutscher Verlag, 7–21.

Mayer, K.U. (1991) 'Life Courses in the Welfare State', in W.R. Heinz (ed.) *Theoretical Advances in Life Course Research*, vol. I. Weinheim: Deutscher Studien Verlag.

Mayer, K.U. (1996) 'Structural Constraints on the Life Course', *Human Development* 29, 163–71.

Mayer, K.U. and Huinink, J. (1990) 'Age, Period and Cohort in the Study of The Life-course: A Comparison of Classical A-P-C- Analysis with Event History Analysis or Farewell to Lexis?', in S. Magnusson and L.R. Bergman (eds) *Data Quality in Longitudinal Research*. Cambridge: Cambridge University Press, 211–32.

Mayer, K.U. and Tuma, B.N. (eds) (1990) *Event History Analysis in Life-course Research*. Madison: The University of Wisconsin Press.

McFate, K. (1995) 'Introduction: Western States in the New World Order', in K. McFate, R. Lawson and W.J. Wilson (eds) *Poverty, Inequlity and the Future of Social Policy. Western States in the New World Order*. New York: Russel Sage Foundation, 1–26.

McFate, K., Lawson, R., Wilson, W.J. (eds) (1995) *Poverty, Inequality and the Future of Social Policy. Western States in the New World Order*. New York: Russel Sage Foundation.

Mednick, S.A. and Mednick B. (1984) 'A Brief History of North American Longitudinal Research', in A.M. Sarnoff, M. Harway and K.M. Finello (eds) *Handbook of Longitudinal Research*, vol 1: Birth and Childhood Cohorts. New York: Praeger, 19–21.

Menard, S. (1991) *Longitudinal Research*. London: Sage Publications.

Merz, J. and Rauberger, T.K. (1993) 'Time Use Questions in a Panel Design. Experiences with the German Socio-Economic Panel', in *Time Use Methodology: Toward Consensus*, *Proceedings of the 1992 Conference of the International Association for Time Use Research*, Rome, June 1992. Rome: ISTAT, 207–18.

Micheli, G. and Laffi, S. (1995) *Derive*. Milan: Franco Angeli.

Mingione, E. (1999) 'Foreword. Longitudinal Research: a Bridge between Quantitative and Qualitative Social Research?', in E. Ruspini (ed.) *Longitudinal Analysis: A Bridge between Quantitative and Qualitative Social Research*, Special Issue of *Quality and Quantity* 33, 3, July–August.

Mooijaart, A. (1998) 'Log-linear Models and Markov Modelling of Categorical Longitudinal Data', in C.C.J.H. Bijleveld *et al.* (eds) *Longitudinal Data Analysis. Designs, Models and Methods*. London: Sage Publications, 318–70.

Mueller, W. (1978) *Klassenlage und Lebenslauf*, Habilitationsschrift. Mannheim: University of Mannheim.

Muffels, R.J.A. (2000) 'Dynamics of Poverty and Determinants of Poverty Transitions: Results from the Dutch Socioeconomic Panel', in D. Rose (ed.) *Researching Social and Economic Change. The Uses of Household Panel Studies*, Social Research Today Series. London: Routledge, 165–87.

Muffels, R. and Berghman, J. (1992) 'The incidence and Evolution of Poverty in the 80s', Working Paper of the European Scientific Network on Panel Studies. Colchester: University of Essex. Paper 24.

Negri, N. (1990) 'Introduzione', in N. Negri (ed.) *Povertà in Europa e trasformazione dello Stato sociale*. Milan: Franco Angeli, 11–72.

Nelson, W. (1982) *Applied Life Data Analysis*. New York: John Wiley.

Nesselroade, J.R. and Baltes, P.B. (1979) *Longitudinal Research in the Study of Behavior and Development*. New York: Academic Press.

Neter, J. and Waksberg, J. (1964) 'A Study of Response Errors in Expenditures Data from Households Interviews', *Journal of the American Statistical Association* 59, 18–55.

Norusis, M.J. (1992) *SPSS/PC+ Advanced Statistics. Version 5.0*. Chicago: SPSS Inc.

Office for National Statistics and Centre for Longitudinal Studies (ONS and CLS) (2001) *Longitudinal Study Newsletter* 24, London, June.

Office of National Statistics (1996) *Family Expenditure Survey*. London: HMSO.

Olagnero, M. and Saraceno, C. (1993) *Che vita è. L'uso dei materiali biografici nell'analisi sociologica*. Rome: La Nuova Italia Scientifica.

Olsen, R. (1996) 'Young Carers: Challenging the Facts and Politics of Research into Children and Caring', *Disability and Society* 11,1, 41–54.

Ongaro, F. (1995) 'La transizione allo stato adulto in un approccio di corso di vita. Persistenze e cambiamenti nel comportamento delle coorti', in SIS (Società Italiana di Statistica) *Continuità e discontinuità nei processi demografici. L'Italia nella transizione demografica*. Cosenza: Rubbettino.

Ostrom, C.W. Jr. (1978) *Time Series Analysis: Regression Techniques*, Sage University Paper Series on Quantitative Applications in the Social Sciences (QASS), n. 07–009. London: Sage Publications.

Øyen, E. (ed.) (1990) *Comparative Methodology: Theory and Practice in International Social Research*. London: Sage Publications.

Payne, C., Payne J. and Heath, A. (1994) 'Modelling Trends in Multi-way Tables', in A. Dale and R.B. Davies (eds) *Analysing Social and Political Change. A Casebook of Methods*. London: Sage Publications, 41–74.

Payne, S. (2001) 'Malattia e ruoli femminili: la relazione tra dipendenza economica, responsabilità di cura e povertà', in C. Facchini and E. Ruspini (eds) *Salute e disuguaglianze. Genere, condizioni sociali e corso di vita*. Collana Transizioni e Politiche pubbliche, Milan: Franco Angeli.

Plewis, I. (1994) 'Longitudinal Multilevel Models: Understanding Educational Progress in Relation to Changes in Curriculum Coverage', in R.B. Davies and A. Dale (eds) *Analyzing Social and Political Change. A Casebook of Methods*. London: Sage Publications.

Ragin, C. (1991) *Issues and Alternatives in Comparative Social Research*. Leiden: Brill.

Rajulton, F. (1999) 'Lifehist: Analysis of Life Histories, a State Space Approach', Paper presented at the Workshop on Longitudinal Research in Social Science – a Canadian Focus, Windereme Manor, London, Ontario, Canada, 15–27 October.

Rajulton, F. and Ravanera, Z.R. (2000) 'Theoretical and Analytical Aspects of Longitudinal Research', Paper presented at the Annual Meeting of the Canadian Population Association held in Edmonton, Alberta, 28–30 May.

Rogosa, D.R. (1995) 'Myths and Methods: "Myths about Longitudinal Research" plus Supplemental Questions', in J.M. Gottman (ed.) *The Analysis of Change*. Mahwah, NJ: Lawrence Erlbaum, 3–66.

Rohwer, G. and Pötter, U. (1998) *TDA User's Manual*, Bochum: Ruhr-Universität Bochum, http://www.stat.ruhr-uni-bochum.de/tman.html.

Rohwer, G. and Trappe, H. (1997) 'Describing Life-courses. An Illustration Based on NLSY Data', Paper presented at the POLIS project conference, European University Institute, Florence.

Rose, D. (1993) 'European Household Panel Studies', Working Papers of the ESF Network on Household Panel Studies. Colchester: University of Essex. Paper 45.

Rose, D. (2000) 'Household Panel Studies: An Overview', in D. Rose (ed.) *Researching Social and Economic Change. The Uses of Household Panel Studies*, Social Research Today Series. London: Routledge, 3–35.

Rose, D. and Sullivan, O. (1996) *Introducing Data Analysis for Social Scientists*. Milton Keynes: Open University Press.

Ruggles, P. (1990) *Drawing the Line. Alternative Poverty Measures and their Implications for Public Policy*. Washington DC: The Urban Institute Press.

Ruggles, P. (1991) 'The Survey of Income and Program Participation. Longitudinal Analysis of Federal Survey Data', Joint Economic Committee, US Department of Commerce, Washington DC: Bureau of the Census. Paper 112. http://www.census.gov/dusd/MAB/sipp~1.html. http://www.census.gov/dusd/MAB/wp112.pdf.

Ruspini, E. (1999) 'Longitudinal Research and the Analysis of Social Change', in E. Ruspini (ed.) *Longitudinal Analysis: A Bridge between Quantitative and Qualitative Social Research*, Special Issue of *Quality and Quantity* 33, 3, July–August.

Ruspini, E. (2000a) 'Longitudinal Research in the Social Sciences', *Social Research Update*, Issue 28. Guildford: University of Surrey.

Ruspini, E. (2000b) *L'altra metà della povertà. Uno studio sulla povertà femminile in Germania e Gran Bretagna*. Rome: Carocci Editore.

Ryder, N.B. (1965) 'The Cohort as a Concept in the Study of Social Change', *American Sociological Review* 30, 843–61.

Sandefur, G.D. and Tuma, N.B. (1987) 'How Data Type Affects Conclusions about Individual Mobility', *Social Science Research* 16, 301–28.

Sankoff, D. and Kruskal, J.B. (eds) (1983) *Time Warps, String Edits and Macromolecules. The Theory and Practice of Sequence Comparison*. Reading, MA: Addison-Wesley.

Saraceno, C. (ed.) (1986) *Età e corso della vita*. Bologna: Il Mulino.

Schaber, G. and Schmaus, G. (1996) 'The Longitudinal and Comparative Research In One: From Patch-Work and Guess-Work to Comparable Data Sets for Comparative Analysis', Paper presented at the Fourth International Conference 'Social Science Methodology'. Colchester: University of Essex.

Schaber, G., Schmaus, G. and Wagner, G.G. (1992) 'Building up an International Comparative Panel Database. The PACO Project', Working Papers of the ESF Network on Household Panel Studies. Colchester: University of Essex. Paper 29.

Scherer, S. (1999) 'Entering the Labour Market: Gender Specific Ways and Patterns', Paper presented at Workshop: 'Beyond the Feminisation of Poverty, Women's Poverty, Defamilialisation and Gender Relationships', University of Padua, November.

Scherer, S. (2000) 'Assetti istituzionali e differenze di genere nell'accesso al mercato del lavoro', in F. Bimbi and E. Ruspini (eds) *Povertà delle donne e trasformazione dei rapporti di genere. Inchiesta* 128, April–June, 75–84.

Schizzerotto, A. (1999) 'L'indagine ILFI sulle famiglie italiane', *Newsletter University of Trento* 16, University of Trento. http://www.unitn.it/unitn/numero16/indagine.html.

Schmaus, G. and Riebschlager, M. (1995) *PACO User Guide: Comparative Research on Household Panel Studies*, PACO Document No. 9. Luxembourg: CEPS/INSTEAD.

Schwarz, N. (1990) 'Assessing Frequency Reports of Mundane Behaviors: Contributions of Cognitive Psychology to Questionnaire Construction', in C. Hendrick and M.S. Clark (eds) *Review of Personality and Social Psychology*, vol. 11: Research Methods in Personality and Social Psychology. Beverly Hills: Sage Publications, 98–119.

Schwarz, N. (1996) 'Survey Research: Collecting Data by Asking Questions', in G.R. Semin and K. Fiedler (eds) *Applied Social Psychology*. London: Sage Publications, 65–90.

Schwarz, N. and Sudman, S. (eds) (1994) *Autobiographical Memory and the Validity of Retrospective Reports*. New York: Springer Verlag.

Schwarze,J. (1995) 'Simulating German Income and Social Security Tax Payments Using the GSOEP', Center for Policy Research, The Maxwell School. Syracuse, NY: Syracuse University. Cross-National Studies in Aging, Program Project Paper 19.

Settersten, R.A. and Mayer, K.U. (1997) 'The Measurement of Age, Age Structuring and the Life-course', *Annual Review of Sociology* 23, 233–61.

Shinn, M.W. (1907) *The Development of the Senses in the First Three Years of Life*. Berkeley: University of California.

Silberstein, A.R. and Jacobs, C.A. (1989) 'Symptoms of Repeated Interview Effects in the Consumer Expenditure Interview Survey', in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds) *Panel Surveys*. New York: John Wiley, 289–303.

Singh, C. (1995) *A Comparative Analysis of Attrition in Household Panel Studies*, Comparative Research on Household Panel Studies (PACO), Document 10. Luxembourg: CEPS/INSTEAD.

Skinner, C. (2000) 'Dealing with Measurement Error in Panel Analysis', in D. Rose (ed.) *Researching Social and Economic Change. The Uses of Household Panel Studies*, Social Research Today Series. London: Routledge, 113–25.

Smelser, N.J. (1981) *Sociology*. Englewood Cliffs: Prentice Hall.

Solga, H. (1998) 'Longitudinal Surveys to Study Changes in Labor-Markets. Advantages and Limits of a Retrospective Design', Paper presented at the International Workshop 'Longitudinal Analysis: A Bridge Between Quantitative and Qualitative Social Research', University of Padua, 14–16 May.

Solon, G. (1989)' The Value of Panel Data in Economic Research', in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds) *Panel Surveys*. New York: John Wiley, 486–96.

Sontag, L.W. (1971) 'The History of Longitudinal Research: Implications for the Future', *Child Development* 42, 987–1002.

Statistics Canada (1995) *Labour and Income Dynamics*, vol. 4, n. 1. Ottawa: Statistics Canada.

Stevens, S.S. (1946) 'On the Theory of Scales of Measurement', *Science* 103, 677–80.

Stevens, S.S. (1951) 'Mathematics, Measurement, and Psychophysics', in S.S. Stevens (ed.) *Handbook of Experimental Psychology*. New York: John Wiley.

Stone, A. (1991) 'Gatekeeping Experts and the Control of Status Passages', in W.R. Heinz (ed.) *The Life Course and Social Change. Comparative Perspective*, vol. II. Weinheim: Deutscher Studien Verlag.

Stouffer, S.A. (1950) 'Some Observations on Study Design', *American Journal Of Sociology* 55, 355–61.

Sudman, S. and Bradburn, N.A. (1982) *Asking Questions*, The Jossey-Bass Series in Social and Behavioral Sciences. San Francisco and London: Jossey-Bass Publishers.

Sudman, S. and Ferber, R. (1979) *Consumer Panels*. Chicago: American Marketing Association.

Svalastoga, K. (1959) *Prestige, Class and Mobility*. London: Heinemann.

Tåhlin, M. (1990) 'Politics, Dynamics and Individualism. The Swedish Approach to Level of Living Research', *Social Indicators Research* 22, 155–80.
Tanner, J.M. (1961) *Education and Physical Growth*, University of London Press.
Tanner, J.M. (1962) *Growth at Adolescence*, 2nd edn. Oxford: Blackwell.
Tanner, J.M. (1963) 'The Regulation of Human Growth', *Child Development* 34, 817–47.
Taris, T.W. (2000) *A Primer in Longitudinal Data Analysis*. London: Sage Publications.
Taylor, M., Brice, J. and Buck, N. (eds) (1995) *BHPS User Manual. Introduction, Technical Report and Appendices*, vol. A. Colchester: University of Essex.
Terman, L.M. (1925) *Genetic Studies of Genius, vol. 1: Mental and Physical Traits of a Thousand Gifted Children*. Stanford: Stanford University Press.
Terman, L.M. (1939) 'The Vocational Success of Intellectually Gifted Individuals', *School and Society* 49, 65–73.
Terman, L.M. and Oden, M.H. (1947) *Genetic Studies of Genius, vol. 4: The Gifted Child Grows Up: Twenty-five Years' Follow-up of a Superior Group*. Stanford: Stanford University Press.
Terman, L.M. and Oden, M.H. (1959) *Genetic Studies of Genius, vol. 4: The Gifted Group at Mid-Life: Thirty-five Years' Follow-up of the Superior Child*. Stanford: Stanford University Press.
Thomas, W.I. and Znaniecki F. (1918–20) *The Polish Peasant in Europe and America*. Boston: Gorham Press.
Torelli, N. and Trivellato, U. (1993) 'Modelling Inaccuracies in Job-search Duration Data', *Journal of Econometrics* 59, 187–211.
Touraine, A. (1974) *The Post-Industrial Society*. London: Wildwood.
Traugott, M.W. and Katosh, J.P. (1979) 'Response Validity in Surveys of Voter Behavior', *Public Opinion Quarterly* 43, 359–77.
Trivellato, U. (1999) 'Issues in the Design and Analysis of Panel Studies: A Cursory Review', in E. Ruspini (ed.) *Longitudinal Analysis: A Bridge between Quantitative and Qualitative Social Research*, Special Issue of *Quality and Quantity* 33, 3, July–August.
Tuddenham, R.D. and Snyder, M.M. (1954) 'Physical Growth of California Boys and Girls from Birth to 18 Years', *Child Development* I, 2. Berkeley: University of California Press, 183–364.
Tuma, N.B. (1994) 'Event History Analysis, in A. Dale and R.B. Davies (eds) *Analysing Social and Political Change. A Casebook of Methods*. London: Sage Publications, 138–66.
Tuma, N.B. and Hannah, M.T. (1979) 'Dynamic Analysis of Event History Analysis', *American Journal of Sociology* 84, 820–54.
Tuma, N.B. and Hannah, M.T. (1984) *Social Dynamics. Models and Methods*. Orlando and London: Academic Press.
Uncles, M.D. (ed.) (1988) *Longitudinal Data Analysis: Methods and Applications*. London: Pion Limited.
van de Pol, F.J.R. (1989) *Issues of Design and Analysis of Panels*. Amsterdam: Sociometric Research Foundation.
van der Heijden, P.G.M. (1987) *Correspondence Analysis of Longitudinal Categorical Data*. Leiden: DSWO Press.
van der Kamp, L.J.Th. and Bijleveld, C.C.J.H. (1998) *Methodological Issues in Longitudinal Research* in C.C.J.H. Bijleveld *et al.* (eds) *Longitudinal Data Analysis. Designs, Models and Methods*. London: Sage Publications, 1–45.
Verma, V. (1997) 'Data Sources and Access for Comparative Analyses', Paper presented at the 'Advanced Research Workshop on Scientific Information, Dissemination and Access in Russia and Eastern Europe'. Colchester: University of Essex, 15–19 August.

Vermunt, J.K. (1993) 'LEM: Log-linear and Event History Analysis with Missing Data using EM Algorithm', WORC Paper 93 909 015/7. Tilburg, The Netherlands: Tilburg University.

Vermunt, J.K. (1997) *Log-linear and Event History Analysis: A General Approach with Missing Data, Latent Variables, and Unobserved Heterogeneity*. Tilburg, The Netherlands: Tilburg University.

Voges, W. (ed.) (1987) *Methoden der Biographie – und Lebenslaufforschung*, Biographie und Gesellschaft, vol. 1. Opladen: Leske and Budrich.

Vogt, P.W. (1998) *Dictionary of Statistics & Methodology. A Nontechnical Guide for the Social Sciences*, 2nd edn. Newbury Park, CA: Sage Publications.

von Eye, A. and Niedermeier, K.E. (1999) *Statistical Analysis of Longitudinal Categorical Data in the Social and Behavioral Sciences: An Introduction with Computer Illustrations*, Assessment Systems Corporation. St. Paul, MN: Michigan State University.

Walker, R. (1995) 'The Dynamics of Poverty and Social Exclusion', in G. Room (ed.) *Beyond the Threshold. The Measurement and Analysis of Social Exclusion*. London: The Policy Press.

Walker, R. and Ashworth, K. (1994) *Poverty Dynamics: Issues and Examples*. Aldershot: Avebury.

Walker, R. and Leisering, L. (1998) 'New Tools: Towards a Dynamic Science of Modern Society', in L. Leisering and R. Walker (eds) *The Dynamics of Modern Society*. Bristol: The Policy Press, 17–33.

Wall, W.D. and Williams, H.L. (1970) *Longitudinal Studies and the Social Sciences*. London: Heinemann.

Waterman, M.S. (1995) *Introduction to Computational Biology: Maps, Sequences and Genomes*, University of Southern California. Los Angeles: Chapman & Hall.

Waterton, J. and Lievesley, D. (1989) 'Evidence of Conditioning Effects in the British Attitudes Panel Survey', in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds) *Panel Surveys*. New York: John Wiley, 319–39.

Watson, D. (1998) 'The 1997 Wave of The Living in Ireland Survey: Response Rates and Sample Weights'. Dublin: The Economic and Social Research Institute and Combat Poverty Agency. Paper.

White, H.C. (1970) *Chains of Opportunity*. Cambridge: Harvard University Press.

Winkels, J.W. and Withers S.D. (2000) 'Panel Attrition', in D. Rose (ed.) *Researching Social and Economic Change. The Uses of Household Panel Studies*, Social Research Today Series. London: Routledge, 79–95.

Wiswede, G. and Kutsch, T. (1978) *Sozialer Wandel*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Wright Mills, C. (1959) *The Sociological Imagination*. New York: Oxford University Press.

Yamaguchi, K. (1991) *Event History Analysis*, Applied Social Research Methods Series, vol. 28. London: Sage Publications.

Yfantopulos, J. (1993) 'Underground Economy and Panel Studies in South Europe. Methodological and Policy Issues', Working Paper of the European Scientific Network on Household Panel Studies. Colchester: University of Essex. Paper 80.

Zajczyk, F. (1996) *Fonti per le statistiche sociali*, vol. 7, Collana Metodologia delle Scienze Umane. Milan: Franco Angeli.

Zazzo, R. (1967) 'Diversité, realité, et mirages de la méthode longitudinale: Rapport introducif au symposium des études longitudinales', *Enfance* 2, 20, 131–6.

Index